Modelling spread of Bluetongue and other vector borne diseases in Denmark and evaluation of intervention strategies

Græsbøll, Kaare

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Modelling spread of Bluetongue and other vector borne diseases in Denmark and evaluation of intervention strategies

Kaare Græsbøll

Supervisors:
Lasse Engbo Christiansen
Claes Enøe
René Bødker

Kongens Lyngby 2012
IMM-PHD-2012-285

Contact:

Kaare Græsbøll	kagr@dtu.dk
Lasse E. Christiansen	laec@dtu.dk
Claes Enøe	clen@vet.dtu.dk
René Bødker	rebo@vet.dtu.dk

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

IMM-PHD: ISSN 0909-3192

The photograph on title page depicting the author is taken by Carsten Kirkeby

Summary

The main outcome of this PhD project is a generic model for non-contagious infectious vector-borne disease spread by one vector species between up to two species of hosts distributed on farms and pasture. The model features a within-herd model of disease, combined with a triple movement kernel that describes spread of disease using vectors or hosts as agents of the spread.

The model is run with bluetongue as the primary case study, and it is demonstrated how an epidemic outbreak of bluetongue 8 in Denmark is sensitive to the use of pasture, climate, vaccination, vector abundance, and flying parameters.

In constructing a more process oriented agent-based approach to spread modeling new parameters describing vector behavior were introduced. When these vector flying parameters have been quantified by experiments, this model can be implemented on areas naïve to the modeled disease with a high predictive power.

Furthermore this PhD has provided a new method of estimating the effect of light traps, which can estimate the additive effect of closely placed traps, and determine trap range of individual traps based on emitted light intensity. Moreover there has been devised a method to sample in time which maximizes information about time dependence and is robust to changes.

Resumé

Dette PhD projekts primære resultat er en generisk agent-baseret model til at modellere ikke-smitsomme infektiøse vektor-bårne sygdomme spredt af en vektor mellem op til to arter af værter fordelt på gårde og græsarealer. Modellen beskriver sygdomsspredning mellem arter i spatielt afgrænsede flokke kombineret med en tredobbelt spredningskerne, der beskriver spatiel spredning af sygdomme ved hjælp af vektorer eller værter.

Hovedformålet med projektet var at undersøge hvorledes sygdommen bluetongue vil påvirke den danske kvægbestand, og det bliver demonstreret, hvordan et epidemisk udbrud af bluetongue 8 i Danmark er påvirket af brugen af græsarealer, klima, vaccination, antal af vektorer og deres flyveparametre.

For at skabe en mere proces orienteret og agentbaseret tilgang til sygdomsspredning var det nødvendigt at introducere nye parametre for at beskrive vektorernes flyvemønstre. Modellens prædiktive effekt vil kunne forbedres med bedre estimater af disse parametre, og dermed vil den være særdeles anvendelig i geografiske områder der hidtil har været uberørt af sygdom.

Udover spredningsmodellen præsenterer denne PhD en ny metode til at beregne effekten af lysfælder, som kan estimere hvorledes tæt placerede lysfælder interagerer, og over hvor lang afstand en fælde tiltrækker vektorer. Derudover er der blevet udarbejdet en metode til at designe prøvetagning over tid, som maksimerer information om tidsafhængighed og er robust over for ændringer i prøvetagningsplanen.

Preface

This thesis was prepared at department of Informatics & Mathematical Modelling in collaboration with the Danish National Veterinary Institute, both of the Technical University of Denmark in partial fulfillment of the requirements for acquiring the Ph.D. degree in engineering.

The thesis deals with modeling spread of bluetongue and other vector borne diseases in Denmark and the evaluation of intervention strategies.

The thesis consists of a summary report, a collection of four research papers written during the period 2009–2012, and elsewhere published, and a technical report describing the code.

Lyngby, November 2012
(Slightly edited before final print May 2013)

Kaare Græsbøll

Papers included in the thesis

[A] Kaare Græsbøll, René Bødker, Claes Enøe, Lasse E. Christiansen. Simulating spread of Bluetongue Virus by flying vectors between hosts on pasture. *Scientific Reports* **2**, (2012).

[B] Kaare Græsbøll, Thomas Sumner, Claes Enøe, Lasse E. Christiansen, Simon Gubbins. A comparison of dynamics in two models for the spread of a vector borne disease. *Scientific Reports*. Submitted.

[C] Kaare Græsbøll, Lasse E. Christiansen. Scheduling sampling to maximize information about time dependence in experiments with limited resources. *Biological Rhythm Research*, 1–11 (2012).

[D] Carsten Kirkeby[†], Kaare Græsbøll[†], Anders Stockmarr, Lasse E. Christiansen, René Bødker. The range of attraction for light traps catching biting midges (Diptera: Ceratopogonidae: *Culicoides*) vectors. *Parasites & Vectors* **6**(1), 67 (2013). [†]Co-authors.

Acknowledgements

First of all I would like to acknowledge the dedicated and inspiring collaboration with my three supervisors: my two official supervisors Claes Enøe and Lasse Engbo Christiansen, and my unofficial supervisor René Bødker. It has been like having three wise uncles, who would always take the time to discuss whatever was on my mind.

I have also had much use of and great discussions with my twinned PhD buddy Carsten Kirkeby. I would like to appreciate the very positive working environment at both the epidemiology section of the Danish National Veterinary Institute and the statistics section of DTU Informatics. A special thank you to all the administrative staff, who most of the time have been downright helpful, especially our department secretary Janne Kofoed Lassen who always knows the best and most efficient ways through the system.

During this PhD I spent three months at the Pirbright Institute, UK, where I thoroughly enjoyed British hospitality by both the mathematical biology and the entomology section. I especially want to thank Tom Sumner and my supervisor at Pirbright Simon Gubbins for being engaged and great partners in my project.

A special thank you to the Knowledge center for Agriculture, Cattle for providing funding and data to the project. Also thank you to the Danish Meteorological Institute for providing temperature data, and Aarhus University for the pasture data.

I will also like to say thank you to the editors and the anonymous reviewers of the submitted and published articles for helping to sharpen the focus of the results.

Contents

Chapter 1

Background

In 2008-2009 the Danish cattle industry spent no less than 5 million Euros on a vaccination campaign to prevent a widespread epidemic of Bluetongue Virus (BTV). The total number of discovered cases of bluetongue in Denmark since first discovered in 2007 is 16 (www.oie.int). No widespread epidemic of BTV has occurred in Denmark or the surrounding countries since 2008 despite that e.g. Norway and Poland did not vaccinate against the disease. The decision to vaccinate cattle, sheep, and goats in Denmark was based on the large outbreak in predominantly the Netherlands, France, and Germany, which for the Netherlands alone had a total cost of approximately 200 million Euros in 2006-2007 [1]. A number of published articles mention a warmer climate as the reason for the incursion of bluetongue in northern Europe [2–5] and consequently it was later speculated that a large scale epidemic would not be possible in the Nordic countries due to a colder climate. This PhD project investigates the possibilities of a BTV epidemic in Denmark, and should such occur what are some possible intervention strategies against the disease. The model framework build for this investigation was also used to investigate the recent incursion of Schmallenberg Virus (SBV), and can potentially be used to model the spread of any non-contagious vector-borne disease.

1.1 Bluetongue

Bluetongue virus is a disease that primarily affects ruminants [6]. Some reports have reported BTV in lions and other African carnivores, however these findings are most likely due to the consumption of ruminants with BTV [7]. The disease was first described in South Africa by Spreull in 1905 [8,9], and BTV is endemic in most of Africa [9]. In areas where bluetongue is endemic the clinical signs in the native ruminant population are usually very mild, and the disease is of no greater concern [6]. However in many European breeds of especially sheep bluetongue has a severe impact with high morbidity and mortality [6]. Also European cattle breeds will experience high morbidity, and though the mortality is usually low, decrease in milk production and increased abortion rates can be costly if the disease is introduced in a naïve population [9]. Other clinical signs of the disease can be fever, swollen head, drooling, lameness, and the famous swollen blue tongue which gives the disease its name [9].

Bluetongue does not spread directly between ruminants, other than occasional transmission from mother to calf during pregnancy [10], and the odd rare case of direct blood contact e.g. a cow eats the placenta from another cow [11]. The primary route of infection goes through an insect vector, this is for BTV female biting midges of species Diptera: Ceratopogonidae: *Culicoides* (figure 1.1). From here on these shall be referred to as midges. There exist more than 1,000 species of biting midges in the *Culicoides* family [12], adapted to local environment midges can be found in all but very cold climates across the world.

Midges are two-winged insects which are typically 1-3 mm long. They are diurnal and as such usually only active in the hours around dusk and dawn [13]. Some species of midges have preferred species to take blood meals from such as cattle, others will take blood meals from whatever warm blooded animal is available [14,15].

To be able to transmit between hosts, bluetongue virus must be sucked in by a midge biting an infectious ruminant host and the virus need then replicate and pass the gut barrier to reach the saliva glans of the vector to be able to transmit to a new susceptible ruminant host [16]. The time it takes to replicate in the vector is depended on the temperature, and there is a minimum temperature ($\approx 13°C$) under which virus cannot replicate [17], this is especially interesting in the temperate parts of the world.

Bluetongue virus has been discovered in 26 different serotypes [18]. The different serotypes show differentiated affinity to both host and vectors [19–21] and may also differ in the incubation period in the vectors [17]. Many serotypes are

Figure 1.1: Blood-sucking female biting midges. Figure reproduced from Wilson, Darpel & Mellor (2008), PLoS Biology 6(8):e210; http://dx.doi.org/-10.1371/journal.pbio.0060210. Published under a Creative Commons Attribution License (CCAL2.5).

geographically confined and bluetongue is today present on all continents except Antarctica [9] (figure 1.2). Surviving bluetongue usually grants immunity towards the disease, but like many other diseases immunity to one serotype of BTV does not ensure immunity to other serotypes [22].

1.1.1 The 2006-2008 epidemic in temperate Europe

Since 1956 bluetongue has been reported to circulate around the Mediterranean Basin [24]. This was particularly serotypes 1, 4, 9, 10 and 16, which was reported to utilize the midge species *Culicoides imicola* [24]. Because of the relative large temperature difference between the temperate and subtropical zone of Europe, and that *C. imicola* is not found north of the Alps, it was generally considered that bluetongue was not an immediate risk for the northern parts of Europe. However in the beginning of the new century a number of Italian studies showed that bluetongue could be found in *C. obsoletus*, a species of midge that is widely abundant in temperate Europe. These findings came with the warnings:

"The isolation of BTV from the Obsoletus complex should be regarded with some concern because it contains at least one species commonly identified across central and northern Europe, where it is known to attack both livestock and human beings." de Liberato (March 2005) [25]

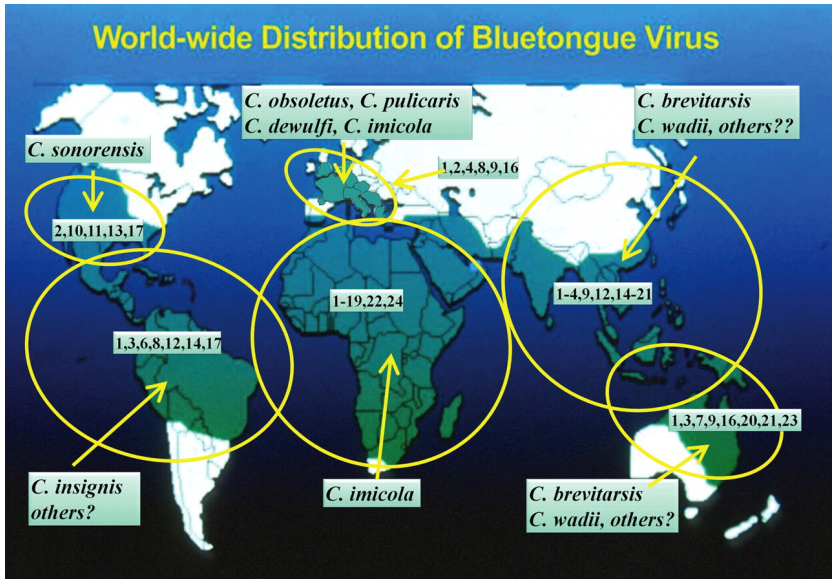

Figure 1.2: The worldwide distribution of bluetongue virus serotypes and the primary Culicoides vectors in different geographical regions denoting six predominant BTV episystems. Figure is from Tabachnick (2010) [23] distributed under open access non-commercial license. First published in journal of experimental biology.

"C obsoletus is distributed from the southern regions of Italy as far as Great Britain, and it is present throughout the year in the southern part of its distribution. As a result, areas which until now have been considered risk-free could be considered at risk of the disease." Savini (July 2005) [26]

Not long after the British concluded that both *C. obsoletus* and *C. pulicaris* midges could in fact transmit the disease as well as the *C. imicola* (Although they had suspected so for some time [27–29]):

"The presence of Culicoides populations in the United Kingdom that are susceptible to BTV infection and are likely to be capable of transmitting the virus, and the total absence of information regarding the susceptibility of other Palaearctic Culicoides species and populations elsewhere in northern Europe, is of real concern." Carpenter (January 2006) [30]

These concerns turned out to be indeed very well founded, because 2006-2008 a massive epidemic of bluetongue serotype 8 (BTV8) spread from the Netherlands into Belgium, Germany, Luxembourg, France, UK, Denmark, Sweden,

and Norway (www.oie.int) affecting more than 90,000 farms [31]. Since then minor incursions have also incurred with serotype 1, 2, 4, 9 and 16 [22]. There have also been raised concerns about the ability of serotypes to cross react in hosts and form more potent serotypes [32].

Moreover bluetongue is not the only virus that the North European midges are capable of transmitting...

1.1.2 Other midge borne diseases

In November 2011 a new disease caused by an arthropod virus was discovered in Schmallenberg, Germany. This virus was sequenced to be closely related to the Akabane and Oropouche viruses [33]: Given that these diseases are midge borne suspicion quickly turned to midges as a vector, and Schmallenberg virus was not long after identified in midges [34]. Schmallenberg presents itself with a short period of clear clinical signs, including fever, decreased milk production, and diarrhea [33]. However, a more serious effect is that if a female ruminant is infected during pregnancy the disease can cause spontaneous abortion and/or misshapen offspring [35]. Schmallenberg virus was confirmed in midges and in cattle in Denmark during summer 2012 [34] (vet.dtu.dk).

When Schmallenberg Virus (SBV) had been coupled to midges the European Food Safety Authority (EFSA) modeled spread of SBV using bluetongue spread models with slightly altered disease parameter values [36]. This was also implemented in the model described in this thesis to asses the potential impact on Denmark. The results of these simulations are presented in section 3.2.1.

Akabane and Oropouche has already been mentioned as midge born viruses in family with the Schmallenberg virus. Akabane is found in Australia, Japan, and Israel, where it is responsible for congenital malformations in ruminants [37]. While Oropuche is a zoonotic disease transmitted between humans and sloths by midges mainly in South America [38]. A few other examples from a long list of ruminant viruses that are midge born are epizootic haemorrhagic disease found in USA and Israel [39] and Shamonda in Japan and Africa [40]. Most midge borne viruses have been shown to dependent on a certain midge species [23, 28, 37–40], and for this reason these diseases are not currently a threat to the Danish industry, but as the case with BTV shows us, diseases may suddenly break this vector species barrier and be able to transmit to new areas.

One midge born disease that has been shown to be transmitted by the same northern European midge species as bluetongue is African Horse Sickness (AHS), which is a disease that affects the horse family with high morbidity and mor-

tality. AHS caused the death of more than 300 horses in Spain in 1987-88, and a widespread epidemic was most likely only prevented by strict movement restrictions and the vaccination of upwards to 300,000 horses, the initial case was most probably an infected zebra imported from Africa to a Safari Park outside Madrid [41, 42].

1.2 Spread models

Models to simulate spread of disease are not a new discipline, but here is not the place for reviewing the entire history of spread models. Rather the most recent spread models of bluetongue shall be presented to clarify to which extent the model in this thesis differs from these models.

The use of the term "spread models" is meant to emphasize that the models include a specific way of handling spatial spread between farms or herds with hosts. Especially the Dutch have published a number of articles on R_0-modeling for BTV based on the Next Generation Matrix (NGM) theory which does not include spatial spread of disease [43, 44]. At the Danish National Veterinary Institute this type of models have also been produced (www.nordrisk.dk), and these models can be of great value when wanting to determine whether epidemic outbreaks are possible at a given location. However, since not including spread these models *a priory* cannot give information about e.g. the rate of spread or how far an epidemic is likely to spread following introduction.

The in-herd dynamics of the model presented in this thesis are based on the works of Gubbins *et al.* and Szmaragd *et al.* [45, 46]. However as presented in article B the spread mechanisms are completely different. The mechanism of the spread model utilized in the model originally proposed by Gubbins *et al.* is based on the concept of transmission kernels, where a function determines the daily probability of transmitting the disease from one location to another solely based on the distance between these two locations. Szmaragd *et al.* demonstrate how this apply for different forms of the transmission kernel [46]. The weakness of transmission kernels is that the simple form encompasses all forms of transmission, which includes all forms of movements of vectors and movements of host. This is very evident from the work of de Koeijer *et al.* (2011) [47] who explores how kernels fitted to different regions of Europe have different shape, which is possibly related to the movement restrictions in the given region. De Koeijer *et al.* also concluded that temperatures influence the transmission probability significantly. Furthermore Pioz *et al.* (2012) [48] presented data that indicates that spread of bluetongue is affected by a number of parameters, where the most significant were elevation, host density, rain, temperature, and vaccination. From

the work of de Koeijer *et al.* and Pioz *et al.* it is evident that the spread of bluetongue is not very well defined by a simple transmission kernel. Now it cannot be claimed that from the start of this project in 2010 it was rejected to make a transmission kernel model based on results published in 2011-2012. The model was build differently to investigate if this would give any effect, and already in October 2010 it could be presented that host distribution influenced the spread of disease (NKVet Symposium on bluetongue, Oslo 2010; figure 2.2). Furthermore Boender presented the initial results from de Koeijer *et al.* (2011) [47] on the GEOVET conference in Sydney 2010 [49], which strengthened the case for continuing along the model path that was originally chosen.

That spread is dependent on meteorological parameters indicates that the spread of the virus is dependent on the abundance of vectors (e.g. humidity in breeding places), or the influence of weather on the movement of vectors (e.g. wind patterns). The next step for modeling spread can then be to explicitly model host and vector movements to better estimate how these are individually affected by e.g. the weather, temperatures, movement restriction etc. and this is done in the model presented in this thesis and in e.g. Turner *et al.* (2012) [50]. Turner *et al.* presents spread of disease by vector movement as a Gaussian transmission kernel that is proportional to the seasonality of vectors, the within-farm prevalence, and number of host on the farm. The Turner model also includes a very detailed representation of how hosts are traded between farms, to demonstrate how this may affect long distance spread of BTV. However Turner *et al.* do not explicitly model the in-herd dynamics on farm, but assume that prevalence is a function of time, and only to a very limited amount temperature. In-herd prevalence can vary much depending on more than a cut-off in temperature, as shall be shown in section 3.2.4. Turner *et al.* (2012) [50] also does not include explicit spread from wind-borne midges.

If a transmission kernel is the simple way of modeling spread then tracing of wind data is the complex way. Here focus shall again be on more recent models, but the reason one is looking at wind in the first place, besides the *a priory* notion that midges are small and fly, is that long distance spread of BTV in the Mediterranean area have been shown to correlate with wind patterns [51–54]. The more resent version of this type of tracking wind plumes are e.g. the model by Hendrickx *et al.* [55] and the NAME model [56] applied onto bluetongue [57, 58] to track possible incursion routes of long distance spread of BTV. The NAME model utilizes meteorological data and knowledge of preferred flying times of midges to emulate midges carried by the wind, so that possible incursion routes or events can be predicted. The NAME model is inherently a particle model to model passive transport from one location to another, and no within-herd modeling is carried out in this model. The model has so far only been used for looking at long distance incursion events between countries, not short or medium spread between herds within countries. This is probably because it

employs very complex 3D modeling of wind patterns which presumably takes a long time to run.

A slightly simpler approach to wind spread is applied by the SWOFT algorithm by Sedda *et al.* (2012) [59] which models wind in 2D, and is therefore capable of model spread between the 2025 infected farms in the 2006 outbreak of BTV8 in Europe. The SWOFT model implements a random kernel, a wind directional kernel, and combinations of the two, to be able to determine which mode of movement is responsible for the transmission of disease. Midges in this model fly between farm locations but farms have a buffer zone with a radius of 750 m to approximate for pasture areas. Midges are set to be able to locate host up to 300 meters away. However this model employs no specific in-herd modeling, it only fits the most likely period from arrival of infectious vectors until clinical signs can be detected, the so called intrinsic incubation period (IIP).

The model in this thesis attempts to compromise between these types of models with an intermediate level of complexity which include explicit in-herd modeling, both active flight and wind carried midges, and host movements.

Chapter 2

Model

From the start it was chosen to do a process oriented stochastic model of spread which includes placing hosts on pasture and midges acting as agents for the disease between herds. The process oriented part means that all parts of the simulation should mimic reality as closely as possible both when modeling within and between herds spread of disease. It was also desirable that the model run fast in order to run sufficient replicates of parameter sets within reasonable time. These two demands are naturally opposing forces, so the model does come with some simplifying assumptions. However, the model is more realistic as a whole compared to other published models of bluetongue spread, given that it includes both explicit in-herd spread of virus and movement of vectors and host.

The model is published in paper A, therefore parts of this chapter are very similar to sections in this paper. However some repetition might be necessary for the ease of read, and the focus shall be on presenting an extended background and further details than presented in the paper.

2.1 Pasture

Because it was desired to model the spread of disease by midges acting as agents for the disease, which includes simulating active local and passive wind flight of

individual infected vectors from cow to cow, it became necessary to distributed hosts onto pasture, since distances between farms and distance between pasture are very different (figure 2.1). If midges where to fly only between farms in the simulation the distance traveled would be much longer than in reality, where they can fly pasture to pasture, and as shown in paper A this greatly affect the size of an epidemic.

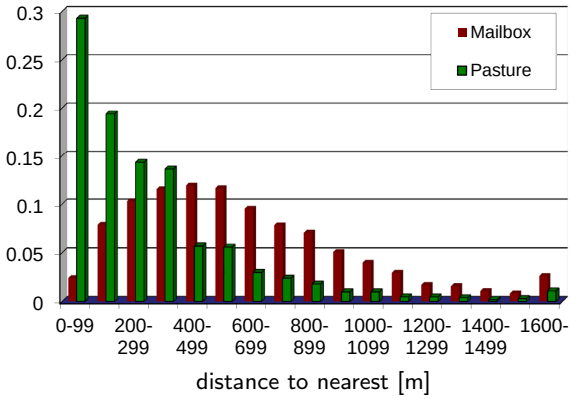

Figure 2.1: Frequency distributions of shortest distances pasture to pasture and mailbox to mailbox belonging to different cattle owners in Denmark. This figure indicates that when midges spread virus from one holding to another the average distance to fly is shorter if cattle is placed on pasture.

Pasture areas were created by dividing land into square grid cells, where the cell side length is an input parameter specified in the program. In each grid cell a number of hosts and vectors can be assigned. The vectors can localize all the hosts which is within the same grid cell, thus the grid size in the simulation is essentially a measure of the vectors' ability to locate host animals. The hosts within one grid cell are designated as a herd, because it is spatially discernable from other hosts. Formally all of Denmark was divided into grid cells, but only cells with pasture or farms contained hosts. Vectors may be present in any grid cell.

The first versions of the model ran with randomly distributed herds, which indicated that the correlation of distances between herds with the size of an epidemic is not trivial (figure 2.2). The figure represent very preliminary data, but the trend that going from 'all on pasture' (short distances between herds) to 'all on farm' (median value of figure 2.1 is app. 550 meters) reduces the numbers of infectious cattle, which is also one of the main conclusions of the final model (paper A and section 3.1.1). What is interesting is that other countries may

have other pasture/farm distributions than Denmark, and figure 2.2 indicates that the efficiency of stabling as an intervention strategy may depends on the host distribution of the affected area..

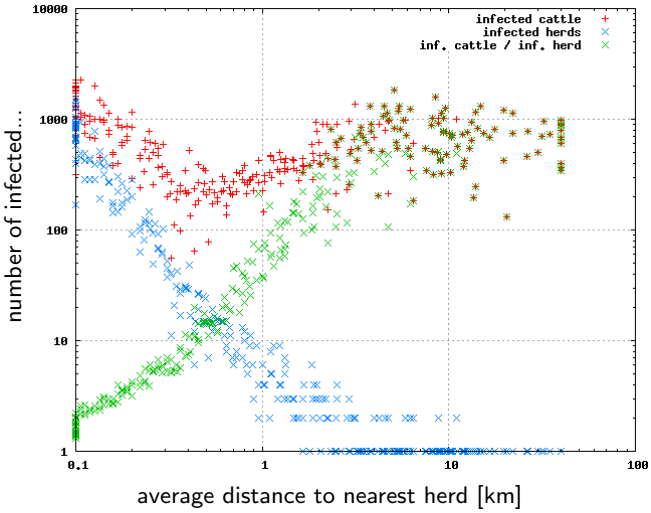

Figure 2.2: The total number of infected cattle (red), infected herds (blue), and infected cattle per infected herd (green) in a simulated epidemic as a function of average distance to nearest herd. These simplified simulations were run in a square box with potential pasture in all grid cells. The simulations have the same number of susceptible cattle, but the spatial distribution differs in that leftmost in the figure there is one cow in each grid cell, while rightmost there is more than 10,000 and up to 40 km to the nearest neighbor containing cattle. These preliminary findings indicated that the farther apart cattle is located the smaller the total size of the epidemic when going from 0 to 500 meters between herds, above 500 meters this trend may be opposite.

To realistically distribute hosts on pasture in Denmark, information from the Danish Central Husbandry Register (CHR) was combined with data from the EU arable land subsidies program linking cattle farmers with potential pasture areas. From this data, 70% of cattle owners were identified as registered landowners of grass-covered areas. This information made it possible to model different 'host on pasture' scenarios ranging from 'all hosts staying inside farm buildings' up to '70% of cattle farmers distributing hosts on pasture'. As it is not exactly known which farmers do and do not put their animals out to grass, it was sampled which farmers had put animals to pasture. Cattle selected for pasture were distributed into the grid cells covering pasture owned by their re-

spective farmers. All animals not put out to pasture were placed in the grid cell containing the coordinate of the owner's farm.

It turns out that the simulations are quite sensitive to the actual distribution of pasture and host (figure 2.3). Which means that it is important to sample this distribution frequently within the simulation, or the output variability might become too small, and the distribution skewed. The model is also quite sensitive to the grid size chosen (as presented in paper A), but as will be discussed in section 2.4 there is no good estimation of the value of this parameter. However in the supplementary material in paper B the grid size is fitted using the data from the 2007 outbreak of bluetongue in England. The grid size applied is typically between 100 and 500 meters.

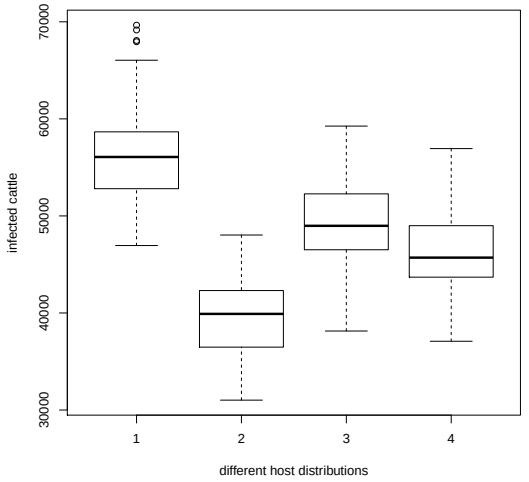

Figure 2.3: Final number of infected cattle in a simulation given four different host distributions (30 replicates each). In the early test phase the model ran for speed considerations with only one host distribution. However, as seen from the figure the median of the final number of infected cattle differs significantly between different host distributions, therefore later models sampled between a number of different host distributions, and work on a module to efficiently sample a new host distribution for each iteration is almost complete.

2.2 Within-herd

Having split Denmark into little squares (referred to as grid cells), the attention now turns to what happens in these.

If a grid cell contains hosts, these hosts are then defined as a herd. This definition identifies spatially discernable hosts, while a holding refers to all animals owned by one farmer, and as such can contain multiple herds. Within each herd the progress of the virus (if present) is modeled by an expanded Susceptible Exposed Infectious Recovered (SEIR) model for the hosts, while vectors are described using an extended Susceptible Exposed Infectious (SEI) model (figure 2.4). This framework is based on the model by Gubbins *et al.* and Szmaragd *et al.* [45, 46]. The extended framework ensures that the model emulates virus incubation time more realistically compared with the non-extended approach. The SEI/SEIR model uses rates which assign a probability of movement of an individual vector/host from one state to the next within one time step. In an unextended SEIR model ($k = 1$) the number of midges becoming infectious per time step is highest on the time step immediately after obtaining the virus, which in the case of bluetongue (and many other vector borne diseases) is impossible by nature (typical values for time to first infectious midge range from 5 to 30 days [17]), especially in areas with low temperature resulting in a long incubation period in the vectors (called the Extrinsic Incubation Period or EIP). Therefore the exposed states of vectors are modeled using extra stages to better emulate the incubation period. These extra stages brings the frequency distribution of vectors completing the incubation period from an exponential distribution to an Erlang (Gamma) distribution [17]. When adding a daily risk of dying to the system, the frequency of vectors completing the incubation period becomes a phase-type distribution [60]. We also use an extended approach on the infectious period of the host to better describe the viremic period.

An example of where it is also important to consider the number of stages is the Next Generation Matrices (NGM) previously mentioned in section 1.2. NGMs can easily be adopted to many different transmissible diseases (See e.g. the PhD thesis by Nienke Hartemink [44] or [43]), but it should be noted that this method is essentially a coupled rate model, and as with other rate models, this model has, if no extra stages is introduced, exponential waiting distribution for all species, and not taking this into account can overestimate R_0 values by as much as a factor of ten (unpublished results and implicitly in e.g. [61]).

As mentioned previously the model is based on the model by Gubbins *et al.* and Szmaragd *et al.* [45, 46]. What differs is that there has been introduced an Exposed stage for the hosts, to emulate the waiting time after a host obtains the virus until it becomes infectious. Furthermore the transmission probability can

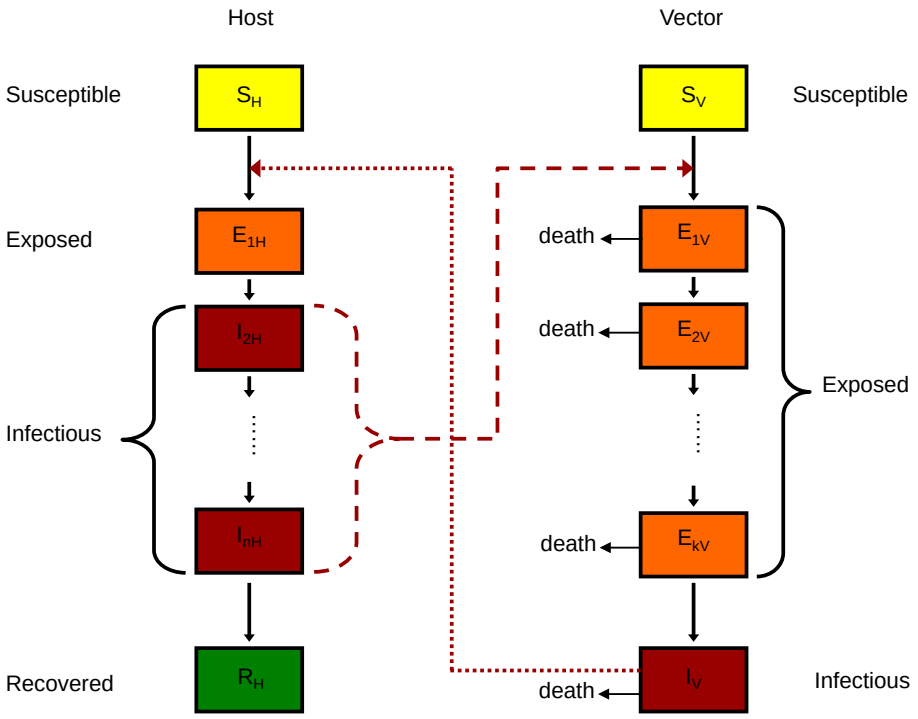

Figure 2.4: The SEIR/SEI model for within-herd disease dynamic. The viraemia of the hosts is described by an extended SEIR model, and the vectors are described by an extended SEI model. All movements between the stages in the model are governed by the probabilities listed in tables A.1-A.2 in paper A. Bluetongue is a non-contagious disease and therefore transmission of virus is modelled through the bite of the vectors, as indicated by the red arrows.

be set proportional with the virus titre in the blood [62] so that the probability of infecting a vector $S_V \rightarrow E_{1V}$ goes from equation (2.1) to (2.2):

$$\frac{a\beta}{H} \sum_{1}^{5} I_{iH} \tag{2.1}$$

$$\frac{a\beta}{H} (0.5 \cdot I_{2H} + 2.0 \cdot I_{3H} + 2.0 \cdot I_{4H} + 0.5 \cdot I_{5H}) \tag{2.2}$$

Where a is the biting rate of the vectors, β is the transmission probability from host to vector, H is the number of hosts in the herd, and I_{iH} is the number of infectious hosts in stage i. The value of the parameters are listed in table A.1 (p. 66). Notice that the sum of scaling factors of I_{iH} both sum to five, so that the total force of infection is the same, it is just differently distributed in time.

2.3 Spatial spread

When virus moves between grid cells or -between herds- it is referred to as spatial spread. In this model the spread is not defined by a transmission kernel but by movement kernels. The difference being that transmission kernels determine the probability of a transmission of virus between holdings, while movement kernels determines the probability of a movement of vectors or hosts. To exemplify the difference: a newly exposed midge might move from one herd to the neighbor herd, but before becoming infectious flies further on without transmitting disease. The length of the EIP depends on temperature (35 days at a daily mean temperature of 15°C, 8 days at 20°C, see table A.1), and therefore the length a midge fly before delivering infectious bites depends on temperature. Furthermore midges may spend time flying between areas with no hosts, and the chance to locate hosts depends on the local distribution of hosts. Both examples on how a process oriented movement kernel might better replicate a transmission event, which is dependent on multiple factors which may change over the season or with the agricultural landscape, compared to a simple transmission kernel, which is an average in time and space over previous transmission events.

Simulated vectors in this model are "free" to disperse in a series of jumps before transmitting the disease, as opposed to traditional kernel models where transmission events are determined by the kernel. The movement of vectors in the model is done by two Gaussian kernels to emulate an active random flight, and a passive spread by wind (figure 2.5, sections 2.3.1 and 2.3.2). Free dispersal in combination with splitting vector movements into two separate Gaussians makes it possible for transmission events to occur as observed in real epidemics

both short, long, and fat tails as dependent by regional farming practices or policies.

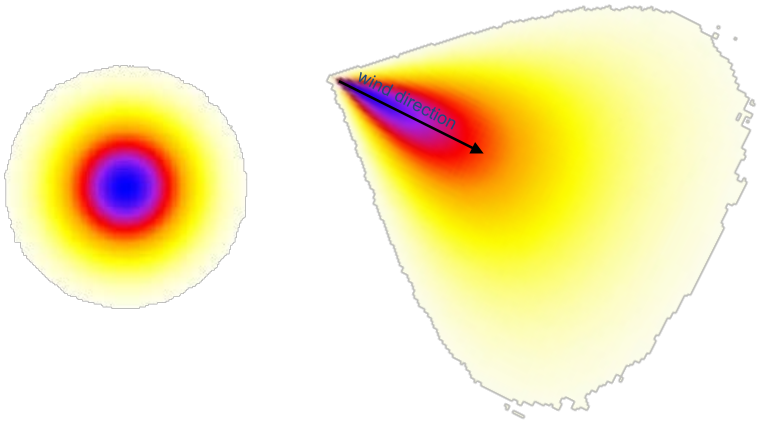

Figure 2.5: The Gaussian kernels that determine the movement of vectors. (Left) probability density when vectors are moving by active flight to neighboring areas (arbitrary length scale, darker colors are equivalent to higher probability, from equation 2.4). (Right) probability density when vectors are carried by wind (arbitrary length scale, from equations 2.5 and 2.6).

Both active (sec. 2.3.1) and passive vector movement (sec. 2.3.2) are applied on each individual vector in each grid cell. However the probability of movement is dependent on whether there are hosts present in the grid cell or not. Therefore two parameters are needed to describe this situation: (i) the probability of flying when there are no hosts in an area; (ii) the probability of flying when there are hosts in an area.

Host movements are also included and described by a movement kernel as presented in section 2.3.3.

2.3.1 Active flight

Within each grid cell it is assumed that the vectors can locate the host animals with 100% certainty. Beyond the size scale of the grid, vectors are assumed to fly a random walk pattern to locate hosts. A random walk can be described by a Gaussian spread kernel, and the flight distance of a vector, d_L, can thus be drawn using the normal distribution in equation (2.3) as the frequency distribution, equation (2.4).

$$f_G(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2.3}$$

$$f(d_L) = 2\, f_G\left(d_L; 0, \frac{d_{L50}}{P_{50}}\right) \tag{2.4}$$

Where the constant $P_{50} = 0.675$ scales so that d_{L50} becomes the radial distance 50% of the vectors will fly within, which shall be later referred to as the active local flight range. The factor of two normalizes the distribution given that distances cannot be negative.

One can easily argument for other types of flight patterns of the vectors; one more obvious is Levy flights, which is used in describing movement patterns when organism display search behavior [63]. However Levy flights requires, as do most other more complicated description of movement patterns, more than one parameter value to describe it. Given that the movement of midges is not very well determined, this model opt for the most simple description.

2.3.2 Passive wind spread

For wind-borne flight the distance carried, d_W, is also drawn with the Normal distribution as frequency distribution.

$$f(d_W) = 2\, f_G\left(d_L; 0, \frac{d_{W90}}{P_{90}}\right) \quad , \quad d_{W90} = t_W w_S \tag{2.5}$$

Where d_{W90} is the distance 90% of the vectors are carried within by the wind. This is a product of the wind speed, w_S, and the time vectors spend in the wind, t_W. Given that and $P_{90} = 1.645$.

The direction of the active flight is drawn from a uniform distribution. While for wind dispersal the direction, θ_W, is distributed as a Gaussian centered on the wind direction, w_θ.

$$f(\theta_W) = w_\theta + f_G\left(\theta_W; 0, \frac{\theta_c}{P_{90}}\right) \tag{2.6}$$

Where $\theta_c = \pi/4$ and P_{90} are given so that 90% of the vectors are within a cone spanning 45° around the wind direction.

The direction of the wind used so far is randomly sampled once daily, and the wind speed is sampled uniformly from 0-5 m/s. The model is structured to accept input of actual wind data. However, it is not expected that actual wind data will influence much on the number of affected herds. But real wind data will probably influence the spatial extent and shape of the outbreak. For UK it was tested what anisotropic sampled wind would do to simulations (figure 2.6).

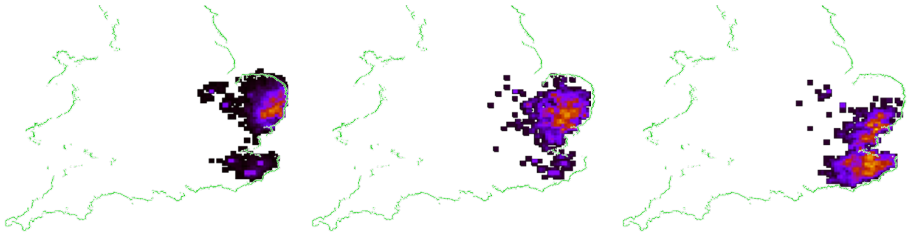

Figure 2.6: Comparison of the spatial extent of a simulated outbreak in the UK when the movement kernel for wind is isotropic (middle) and when the wind is predominantly coming from SW (left) or NE (right). Each sub-figure represents 30 replicates produced as specified in the supplementary material in paper B except for the anisotropic wind. The color represents fraction of simulations where virus becomes present in an area ranging from black to yellow (low to high presence).

In figure 2.6 vectors are assumed to fly with the wind. Whether this is a valid assumption is difficult to establish; Sedda (2012) [59] concluded from the outbreak of BTV8 in northern Europe that midges fly against the wind if they can fly on average 0.5 m/s, but if they only fly 0.13 m/s their results were not significant. In section 2.4 the average speed of midges is estimated to 0.09 m/s from the outbreak in UK.

2.3.3 Host movements

Movement of host animals is modeled using a movement kernel, which states the daily probability of a herd having hosts moved, p_M, and specifies the distance [km] hosts were moved, d_M, should such a move occur. The equation 2.7 was fitted using data extracted from the movement register in Denmark, which tracks all movements of cattle.

$$d_M(x) = 12 \ \log^2\left(\frac{1.01 - x}{1.2}\right) \tag{2.7}$$

Where x is drawn from a uniform distribution on $]0;1]$. Data on movements were from the period October 2006 to September 2007, and therefore this data is not affected by the movement restrictions implemented after the discovery of bluetongue virus in Denmark.

A more process oriented approach to host movements would have been to conduct a network analysis of the Danish cattle movement database to determine nodes and transmission routes. In paper A it was concluded that host movements in Denmark does not contribute greatly to the spread of an epidemic when using the movement kernel, and therefore there should be no greater effect of implementing the full network analysis. However, the movement kernel is an average over host movements, and if bluetongue should affect a holding with a large trade to other holdings this may have a significant effect on the spatial spread.

2.4 Parameter estimation

This new process oriented model requires an evaluation of five new parameters, which have not been evaluated for midges before. Two experiments carried out in 2010 by Carsten Kirkeby at the National Veterinary Institute were performed to establish a first estimate for the five unknown parameters: (i) the area within which midges locate hosts (grid size); (ii) the probability of flying when there are no hosts in an area; (iii) the probability of flying when there are hosts in an area; (iv) the active local flight distance; and (v) the passive flight time in the wind. Unfortunately the data from these experiments have not yet been fully compiled, as they are part of an ongoing PhD project. However, here it shall be briefly mentioned how these data sets could have been used to estimate the parameters.

The first experiment placed midge traps for 25 days around a farm with 700 cattle stratified for distance and direction. Figure 2.7 displays simulated data of a simplified setup using parameters (i)-(iii). Two million midges are distributed in concentric circles around a central herd of cattle. The probability of moving to another circle is depended on the distance to cattle, so that midges outside (i) moves with probability (ii)/2 and midges within (i) moves with (iii)/2, the halves represent that they could move orthogonal to the position of cattle. This model is run until a dynamic equilibrium sets in, and then we sample the number of midges caught in each circle, repeated 25 times. This example is very simplistic and the data is way to nice, but what we would hope of the real data was that a cutoff can be seen to reveal (i), and that the proportion of midges between the sides of the cutoff can be used to derive probabilities (ii) and (iii).

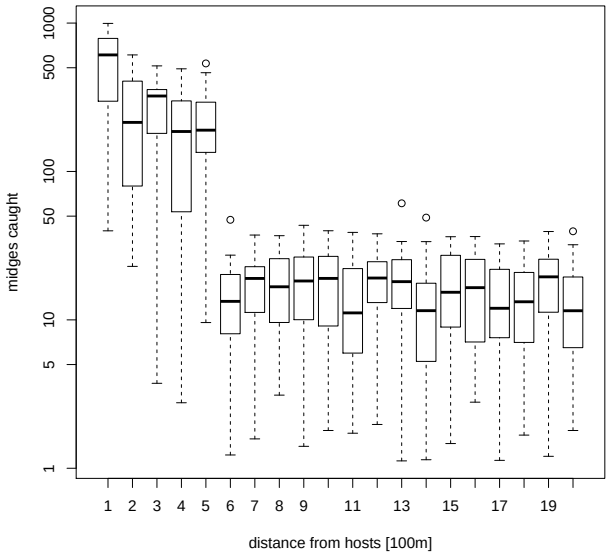

Figure 2.7: Simulated number of midges caught in four traps per distance band as a function of distance from hosts. From this simulated data set three of the unknown flying parameters of the simulation can be fitted.

The second experiment was done with the same setup. On four days midges caught in the traps located next to cattle were marked with dye and released to be hopefully later recaptured. This experiment should be able to produce an estimation of how many midges leave the area with hosts (iii), as well as an estimate of the total number of midges. Furthermore combined with wind data from a mini weather station deployed in the area, it could also give some estimate of how long midges fly on days with and without wind to estimate (iv) and (v), as well as whether they fly with or against the wind. (Note: The results from this study were published after the defence of this PhD thesis. These results indicate that midges may prefer to fly upwind [64].)

Since these parameters were not determined a sensitivity analysis was carried out in paper A, and a maximum likelihood method was used to estimate these parameters from the outbreak in UK in paper B, Supplementary Material SM1. It was there estimated that midges move a median distance of 500 meters per day and/or fly 1.5 hours in the wind. If it was assumed that midges also use 1.5 hours on their active flight this gives an average speed of 0.09 m/s. That midges should only fly 1.5 hours a day is plausible given their diurnal nature, as observed in midge catches by e.g. Sanders *et al.* (2012) [13].

My fellow PhD Carsten Kirkeby did provide some data, and from this we did

a joint publication D to determine the trap range of midge traps, which should aid in the interpretation of future data. Furthermore, some older midge trap data was used as an example in paper C to determine how to do more efficient temporal sampling.

Most of the within-herd parameters were previously determined and this model runs with within-herd parameters similar to Szmaragd (2009) [46] except that the EIP and number of stages to decribe the EIP was better established in Carpenter (2011) [17] and hence used in our work. Also the model uses point estimates of parameter values instead of sampling in a range. However, paper B showed that the models are not sensitive to sampling in the within-herd parameters.

2.4.1 Midge abundance

The primary vector of bluetongue virus in Denmark and the temperate parts of Europe is the female biting midge of species *Culicoides obsoletus* and *C punctatus*. The abundance of the vector is naturally the primary driver of disease spread. However, this is also a parameter that is difficult to predict. Catching midges on a daily basis displays extreme variations from zero one day to thousands the next. The difficulty with catching midges is usually attributed to weather conditions primarily wind and precipitation. Besides the difficulty to catch midges, they also display a seasonal behavior that can be described and has been modeled as between one to five generations per season [65, 66]. Denmark is usually declared vector free from December until April (personal communication with Carsten Kirkeby and www.foedevarestyrelsen.dk). Furthermore, the different midge species have different preferences with regards to breeding sites, which may also lead to differences in the number of generations and timing hereof between species.

In this PhD thesis midge abundance is modeled with basis in the collection of midges and abundance model by Nielsen *et al.* [65] in Egeløkke Lung, located centrally in Denmark. Abundance is described with four peaks in abundance between June 1 and November 7.

2.5 Intervention strategies

The title of the PhD promises an 'evaluation of control strategies' this will include vaccination, housing, and movement restrictions.

Vaccination was implemented on the UK data set (see paper B). Vaccination was modeled by reducing the probability of transmitting virus from host to vector and vice versa. The level of protection was modeled as a linear function that goes from zero to 100% as a function of time from vaccination day to time of full protection. The time to full protection in sheep was set to 14 days, and 60 days in cattle. The time of vaccination for each county was replicated according to the Defra roll-out plan, which was implemented in 2008. This method was similar to the one presented in Szmaragd (2010) [67]. Results from this study is presented in 3.1.1 and paper B.

Vaccination in Denmark was done by testing scenarios with different levels of immunity in the herds. The level of of immunity can be translated to time since a countrywide vaccination as presented in section 3.2.2.

One of the main investigations of paper A was how the fraction of cattle put to pasture influences the size of an epidemic. The results also indicate the opposite: What will be the effect of pulling cattle from pasture into stables, which is equivalent to housing as an intervention strategy.

Restrictions on movement of hosts were included in the code. First of all restriction were implemented trivially by just turning off the movement kernel. Secondly putting a limitation on the kernel so that hosts may only be moved a certain distance from the infected farms, in better concordance with actual movement restrictions imposed by the European Union. However, effects of movement restrictions were not investigated as movement of hosts was not influencing the epidemic significantly.

In the project description for this PhD the use of vector control was also suggested as a possible control mechanism. Currently there are ongoing investigation under the EDENext framework to determine the influence of insect repellant on transmission parameters of BTV for European bluetongue vectors (www.edenext.eu). Carpenter *et al.* (2008) [68] extrapolated from experiments with other midge borne species to conclude that pyrethroids and stabling may aid against an epidemic outbreak. There have been studies that suggest that managerial practices influence the abundance of vectors on the farm [69]. This is primarily in connection with handling of manure, and it is of course advisable to manage a farm in a matter that reduces the presence of potential agents of disease.

2.6 Generalisability

The previous sections have related this model of vector-borne disease spread to bluetongue virus in UK and Denmark. This section will focus on the generic aspects of this model.

That the model has attained such a high level of generality is both an effect of necessity and the process oriented approach. 'Necessity' because there has been no large-scale outbreak of bluetongue in Denmark, and given that there was no data to fit the model to, it was needed to make a model that should in principle work in an area where no transmission kernel could be fitted to. This lead to a process based approach which was then decided to be implemented as flexible as possible.

The current version of the code runs with one vector species and one or two host species. Which vector or host used is in principle only determined by the parameters that is centrally stated in the program. This means that the model can run any non-contagious vector borne disease by only changing a few parameters values. This was demonstrated as the Schmallenberg outbreak was discovered, the model could present results on disease spread a few hours after it was decided to do so (results in section 3.2.1).

The model does require a very detailed input of temperature, pasture, farms and the coupling between these. However, when this is available it can in principle run on any area, with the possible exception of mountainous areas. The abundance of midges are affected by elevation, this is probably mainly due to factors such as temperature, humidity, soil, and increased wind speed [70]. If midge abundance and temperature is present as input data, the model should in principle run, however mountains might also change flight behavior of midges which prompts caution if applying the model to such areas.

The model is structured around a number of modules which handles the different aspects of an epidemic, and it would be possible with relative ease to add modules that handles e.g. contagious disease spread. The functions of these modules are described in appendix E.

2.7 Code

The model is coded in Fortran 90 and it can run in parallel using the open MPI standard. A description of all modules and subroutines in the code can be found in appendix E, including a link to download the code.

Appendix E does in general explain what goes on in the code, and the reasons for the choice of mathematical description in the code have been gone through in the previous sections. Here it shall be briefly reviewed what is done to speed up the run time of the model.

The code is a Markov Chain Monte Carlo style code, which means that each time step is only dependent on the state of the system in the time step before. This is a very standard way of saving on processing power and reducing the information need stored. However this has the drawback that it is not possible to follow a single host or vector over time, as these are handled collectively per grid cell.

All grid cells containing virus are indexed and linked to lists that handle infected animals. This means that the code need not scan across lots of non-infected herds, and all viraemia can be handled non-spatially. The lists are dynamically allocated to save on memory, and they are done so in chunks to increase speed.

The code was unofficially benchmarked in speed against the transmission kernel model by Szmaragd *et al.* [46] in paper B. Where it ran at comparable speed when simulating infection of 30,000 holdings in the UK over one year. When running shorter or with higher vaccination to the effect of fewer infected holdings this model is faster. However with larger outbreaks the model is comparable slower. Both the fast and the slow behavior are due to the explicit modeling of vectors.

Chapter 3

Results

The results from this PhD project encompass model results, and results where the focus have been on how to achieve better input parameter values for this and future models on vector borne diseases. First a summary of the published results, with a focus on how these results relate to each other. Then in section 3.2 some additional results, that are currently not published.

3.1 Results summary

Here follows a summary of the major findings from this PhD project.

3.1.1 Model results

The results presented in this section are primarily from papers A and B.

It was establish in section 2.6 that the model can be adapted to other countries than Denmark, this was testes during a three months stay at the Pirbright Institute (at the time: Institute for Animal Health, Pirbright) in England, where the

model was implemented on the 2007 incursion of bluetongue in UK and compared to the transmission kernel model presented in Szmaragd *et al.* (2010) [67]. The primary results (paper B) from this direct comparison were that our model showed higher sensitivity to midge abundance (the number of infected hosts was more directly proportional to midge abundance) and vaccination (the number of infected hosts was directly proportional to the level of vaccination) compared to the English model (figure B.1 in paper B). It has also a lower variability of output (figures B.1 & B.2) and a more stable initialisation. Furthermore our model reacts better and more biological correct to changes in temperature (figure B.2). From the 2007 English outbreak data of bluetongue it was also possible to fit at least the scale of flying parameters, even though many of them are interdependent (paper B, SM1).

Furthermore, In paper A it was shown that the model is sensitive to temperature, the use of pasture and the five flying parameters already mentioned in section 2.4 (figures A.1-A.4):

Later it shall be shown that temperature is probably not the primary reason that Denmark did not experience a large-scale epidemic (section 3.2.3), However should the average temperature change by one degree the simulations show a change of approximately a factor of five in the number of affected cattle (figure A.3). The temperature dependence in the simulation is caused by the biting rate, the extrinsic incubation period, and the mortality rate all related to the vectors being temperature dependent (table A.1). These three equations reveals that there is a limit to how long increasing temperatures results in larger epidemic. Firstly the expression for the biting rate is only valid up to 41.9°C, and secondly the mortality rate goes past one at 29.4°C. These limits are, however, artificial since they are beyond the data range the expressions were fitted to in the original articles [71, 72]. However before these data limits are reached the increase in temperature will limit BTV as shall be discussed in section 3.2.4. It should also be noted that most parameter values are not determined for the Danish midge species, which is a large source of uncertainty.

The use of pasture is the greatest difference between the model presented in this thesis and other models. The sensitivity analysis on use of pasture is therefore one of the most interesting results in this thesis. The result shows that distributing hosts onto pasture increases the size of an epidemic both in numbers and in area (figures A.1 & A.2). This effect is limited by the ability of vectors to locate hosts, so if vectors can navigate, by whatever sense, directly between farms as distributed in Denmark the pasture effect is significantly lessened.

In section 2.4 it was discussed how data to determine the five undetermined flying parameters might look like. Since the parameters are not determined the result is the sensitivity analysis of the parameters (figure A.3). The sensitivity

analysis shows that there are parameter values which give both under- and over-dispersion, that is the disease is only spread very slowly or virus is spread so much that it thins itself out.

Returning to initialisation: In paper B it is stated that only 2.7% of the simulations performed by the transmission kernel model takes off, compared to 100% in the agent based model. In the UK simulation the model is initiated at seven given locations with five infectious midges. When the model is run on Denmark it is only initiated in one location usually with five or ten midges. As it is quite relevant how many midges are needed to begin an epidemic figure 3.1 display the final size of an epidemic as a function of number of initial infectious midges. This shows that as soon as the number of infectious midges is above three there is little effect on in the total number of affected cattle when increasing the number further. It is interesting that in 32% of simulations one infectious vector will be enough to initiate an epidemic.

The scale of an epidemic is also dependent on the location of introduction (figure 3.2), as an introduction in a cattle dense area will lead to significantly more affected animals than in cattle sparse areas.

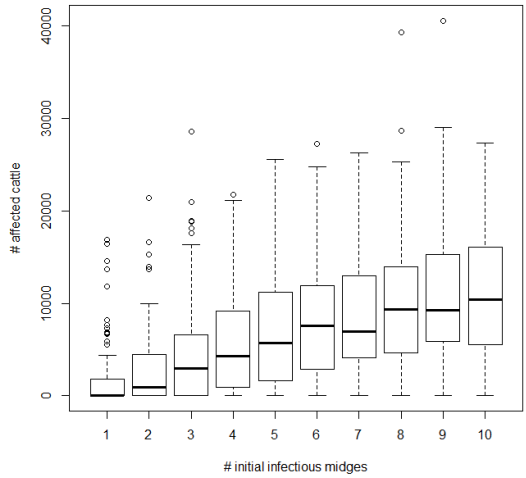

Figure 3.1: Final number of affected cattle in an epidemic as function of number of initial infectious midges arriving in southern Jutland at day 150 in 2008. 100 repetitions with the standard parameter setting. The graph indicates that there is no significant difference in median epidemic size when between 5 and 10 infectious midges initiate an epidemic.

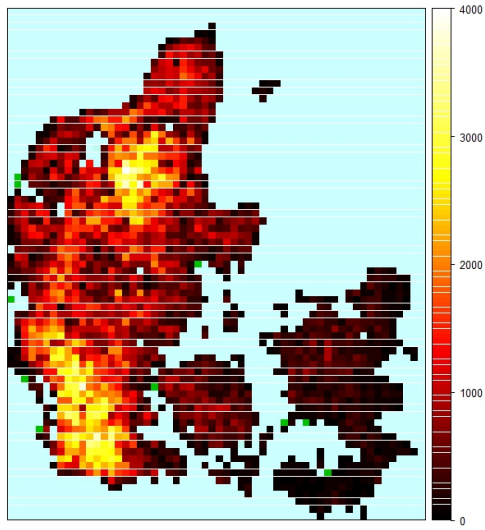

Figure 3.2: The average number of affected hosts as a function of introduction site. Each simulation introduced virus one time and ran to the end of vector season to determine final size of epidemic. Simulations are repeated once for each grid cell containing hosts, and then averaged in pixels of 5 by 5 km. The average final size is strongly correlated to the cattle density in Denmark.

In this thesis all runs of the spread model is, if nothing else stated, done with the parameters listed in table A.1 (page 66), a grid size length of 300 meters and 35% cattle on fields, introducing 5 infectious midges in the location seen as a white square in figure A.2, using temperature data from Denmark in 2008. The reasons for the choice of this set are given partly as the fitting done in paper B, SM1 page 82, the parameter values from table A.1 on page 66 are typically median values estimated from the literature (references found in the table), and 35% cattle on pasture is a guesstimate based on the general knowledge available in the department.

3.1.2 Temporal sampling

As mentioned in the introduction of paper C, we wanted to estimate the biting rate of Danish midge species from previous data (which is the reciprocal of the mentioned egg-laying cycle), since this would increase precision of the within-herd part of the model. However because data was sparse and sampled with random intervals it was not possible to deduce any temporal correlations. A

small investigation into remedy this situation for the future turned out to become a very general and practical tool for sparse temporal sampling, especially in the case when sample times are of risk of being moved.

The tool that was created plans how to place n samples over a period of N time units. The samples will be placed so that the number of observations in every time lag is maximized, to better the significance level of the autocorrelation function. Furthermore the sampling plan is made robust by ensuring that canceled samples can be moved to future time slots without large loss of information. This new method of sample planning enables a prediction of biting rates from simulated data twice as efficient as a random plan, and also significantly better than periodic sample plans. The proposed method to produce a sample plan is shown to be better also in the case of forced movements of sample points.

3.1.3 Attraction range of midge traps

Many of the parameters in the model relating midges are not very well determined (section 2.4), to get better estimates further experiments must be conducted. Many of these experiments will be based on the catch of midges by light traps. In paper D the function of traps was modeled to estimate the trap range. A better understanding of how midges are attracted to light traps will hopefully lead to better estimations of the parameters needed for the spread model. To exemplify: When trying to establish the cutoff within which midges can detect animals, it becomes important to know how far away from their placement traps attract midges.

The principal findings of paper D are that the anisotropic nature of the light from the traps light-tubes must be taken into account (figure 3.3), and that traps placed closely together does not only compete, but also collaborate in the sense that the total light intensity *a priory* increases which attracts midges from farther away and outshine neighboring traps which are not placed as closely. We found that the range of attraction of a single 4W CDC trap was $15\frac{1}{4}$ meters orthogonal to the long side of the light tube.

3.2 Additional results

The model can potentially deliver more results than presented in the articles, and in chapter 5 some of these opportunities are discussed. Below some results are presented, which have not yet fully matured to publication readiness.

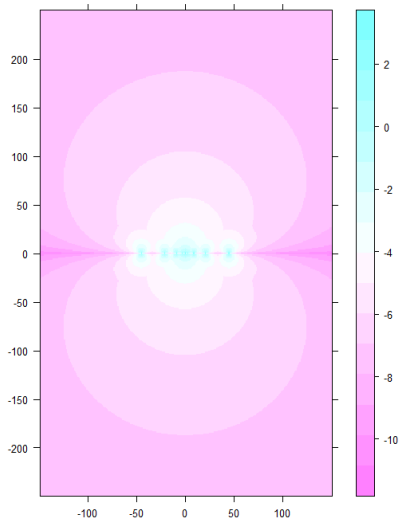

Figure 3.3: Maximum light intensity as perceived by midges around ten traps placed on $y = 0$ as half of the described in setup A (figure D.1 in paper D). The light-tubes in the lamps are all lined up with the x-axis, and the anisotropic properties (more light is emitted perpendicular from the tube than from the end pieces) of the light-tubes then result in visibility being much greater in the y-direction compared to the x-direction.

3.2.1 Schmallenberg

In 2012 Schmallenberg was detected in DK [34]. As mentioned in section 2.6 the model was fed with as much information as was known about Schmallenberg virus on the basis of an EFSA technical rapport released 6 February 2012 [36] to produce the results that will be presented in this section. To initialize these simulations, data from the German Research Institute for Animal Health the Friedrich-Loeffler-Institut (www.fli.bund.de) was used to establish the two nearest infected farms to Denmark (Schmallenberg Virus Karte Stand 21, from 2012-02-21). From this information the most likely risk of presence was simulated as presented in figure 3.4.

When the highest risk of presence was established in figure 3.4, it was investigated what the prevalence of disease would be if present in a given pixel (data was aggregated by area in 5 by 5 km pixels) as in figure 3.5, which show prevalence to be inversely proportional to cattle density. The reason that prevalence is low in cattle dense areas is that it takes time to spread to between cattle, and if the time window where virus can spread is short the disease cannot infect all cattle. Combining these to figures by multiplying them together gives figure 3.6, which indicates the most likely area to detect the disease. Incidently the first midge containing SBV was found in the middle of the hotspot in figure 3.6 [34].

In these simulations the only parameter that is different from the ones used when simulating bluetongue is the infectious period of cattle. Here it is set to 6

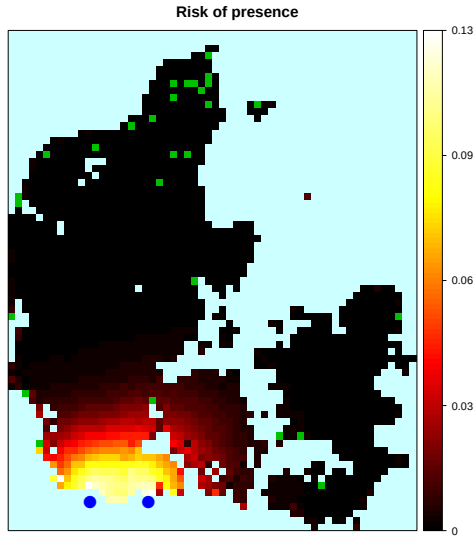

Figure 3.4: Fraction of simulations that show presence of Schmallenberg in 5 by 5 km pixels given two infected German farms at the blue spots (location of DE farms from www.fli.de). There were run 100 replicates with starting date at each day of the year, in total 36500 simulations. Green color is pasture with exactly zero number of simulations showing presence of disease.

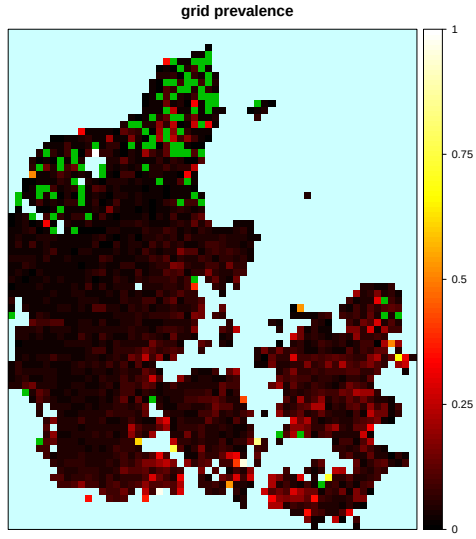

Figure 3.5: The average prevalence of SBV in the host animals per pixel, given SBV occur in this pixel (equivalent to zero prevalence being excluded when calculating average). Simulated as in figure 3.4.

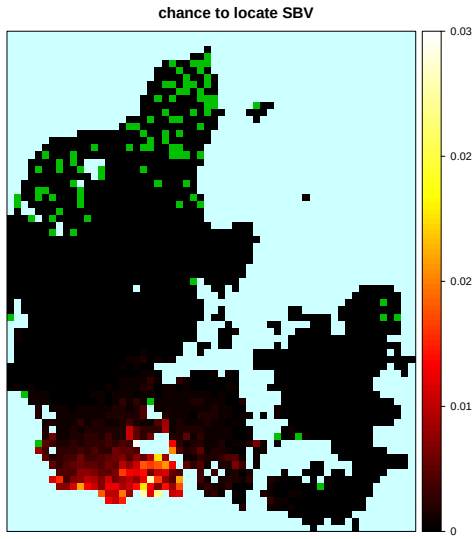

Figure 3.6: Multiplying the results from figures 3.4 (probability of presence) and 3.5 (prevalence if presence) determines the most likely area to detect SBV.

days, the EFSA report [36] estimates a range of 1-6. Which is much lower than for bluetongue which is 20.6 days [73]. Schmallenberg has been detected as far north from Germany as Finland, which suggest that SBV has spread farther (and perhaps faster) than BTV (www.evira.fi). Given that the parameters used to model are not in the favor of SBV something else must be different, which shall be discussed further in chapter 4.

3.2.2 Vaccination in Denmark

Denmark vaccinated approximately 55% of our cattle against BTV8 during the campaign in 2008 (www.landbrugsinfo.dk). To investigate the protection over time, a series of simulations were run to examine how large a fraction of each Danish herd must be susceptible before an outbreak can occur in a worst case scenario (figure 3.7).

If it is assumed that immunity is lifelong, and that the average age of a Danish dairy cow is 60 months (www.landbrugsinfo.dk), then every 10% drop in immunity corresponds to approximately 6 months past a full vaccination of all dairy cattle in Denmark. To achieve a more realistic result in this type of simulation the age distribution at each farm could be extracted from the cattle database,

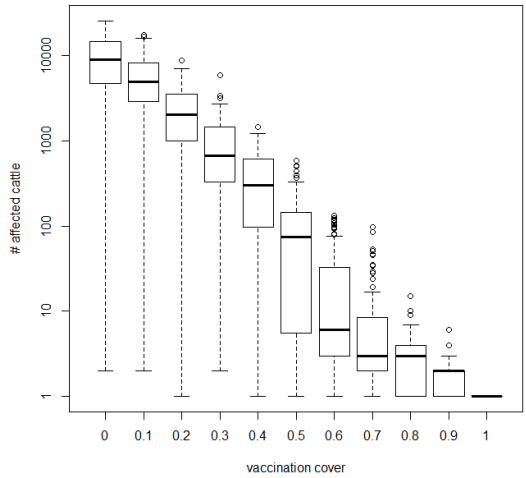

Figure 3.7: Final number of affected animals as a function of intra herd vaccination cover in all herds. 100 repetitions per cover level using the standard setup. One has been added to the data, so that it may be represented using a log-scale. Most noticeable is the drop between 50% and 60% vaccination uptake.

and used to approximate the temporal level of immunity at herd level following a national vaccination.

3.2.3 Temperatures

It has been speculated whether the cold temperatures in the Nordic countries will prevent a wide scale epidemic. If we compare temperatures in the Netherlands, who experienced a massive epidemic in 2007, with the temperatures in Denmark on the year of most outbreaks (2008) they only differ substantially in the very beginning of the Danish vector period (app. day 90-330) (figure 3.8), but as shall be discussed in the discussion (chapter 4) other non-parameterized variables affected by climate factors may influence the epidemic. To compare the effect of temperatures in DK-NL the model was run using only the within-herd part to determine how in-herd prevalence is affected by the difference in seasonal temperature (figure 3.9). This run was done with a constant vector to host ratio, and if shows that if midges are abundant an epidemic in the Netherlands could potentially start one month earlier than in Denmark. In figure 3.11 when an introduction is considered at the worst time possible (day 150), outbreaks in NL and DK would experience the same level of in-herd prevalence of

disease.

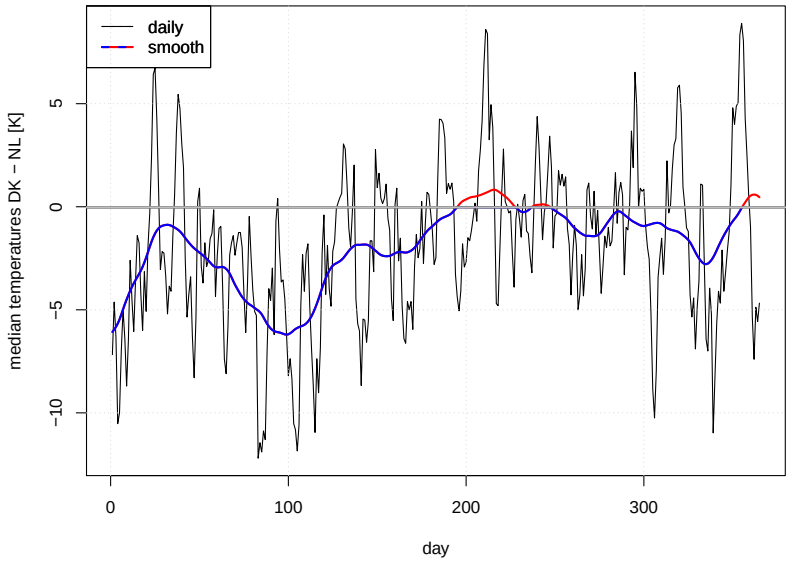

Figure 3.8: Difference of daily mean temperature between Denmark 2008 (66 stations) and the Netherlands 2007 (31 stations, NL data from www.knmi.nl). Within the period of active vectors (app. day 90-330) more than 90% of the period has a temperature difference of less than 3 degrees Celsius. (If comparing year 2008 directly 90% has less than 2 degrees difference, not shown.)

3.2.4 Within-herd prevalence

In section 1.2 it was noticed that some models of bluetongue do not include within-herd models. Turner *et al.* [50] uses a mathematical description, which is partly temperature dependent. While Sedda *et al.* [59] fits an Internal Incubation Period (IIP) of eight days whereafter farms where considered to produce infectious midges. Data from our model suggest that within-herd prevalence may not be homogeneous and also dependents on other parameters than temperature i.e. the vector abundance. In figure 3.10 it is seen how the vector to host ratio in the model will affect the prevalence of infectious hosts as a function of time after introduction. The final seroprevalence (total fraction of hosts affected by disease) of the presented data in figure 3.10 is 20% and 95% (not shown). To investigate this effect further the model was run using only the

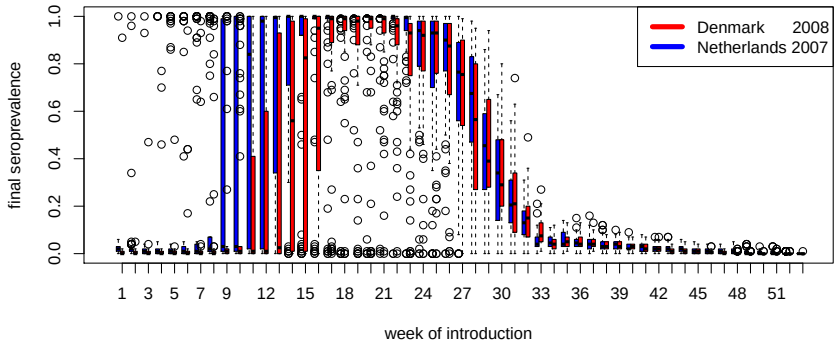

Figure 3.9: Final in-herd sero-prevalence in one farm following introduction of five infectious midges in a given week using temperatures from the Netherlands 2007 and Denmark 2008. Each box represents 70 repetitions with a vector to host ratio of 5000. The simulation is run without spatial spread using only the SEIR/SEI model presented in figure 2.4. The warmer spring in the Netherlands 2007 allows for an earlier onset of epidemic of BTV.

within-herd part to investigate how final seroprevalence depends on temperature, vector abundance, and day of introduction. The figures 3.11 & 3.12 show that temperatures need to be in an optimal area and midge abundance (vector to hosts ratio) needs to exceed a certain threshold for in-herd seroprevalence to become high. The influence of temperature is indirect in figure 3.12 which display final seroprevalence as a function of day of introduction. This figure also depicts that virus must be introduced within a certain time window to reach a high in-herd prevalence.

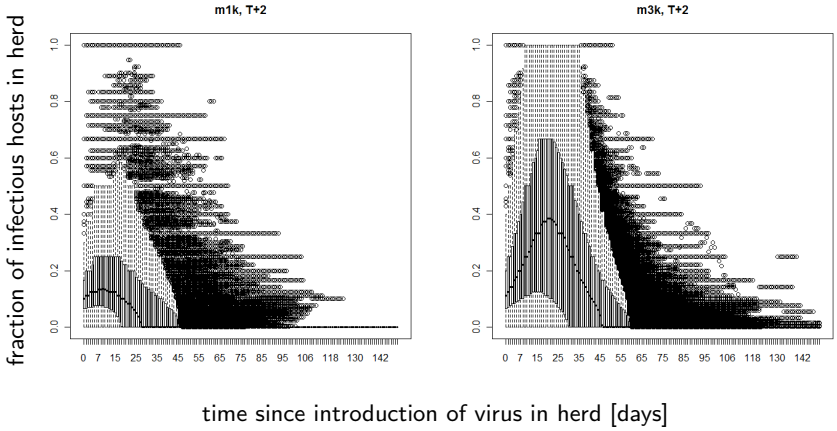

Figure 3.10: In-herd prevalence of infectious hosts as a function of time since infection reached herd, following an introduction of five infectious midges on day 150 in the Southern part of Jutland, Denmark, with subsequent spread. Using temperatures from Denmark 2008 with offset to the temperature of two degrees. Left figure has a vector to host ratio (vth) of 1000, right figure has vth=3000. These results from the full spread model are in good concordance with the results when using only within-herd as in figure 3.11.

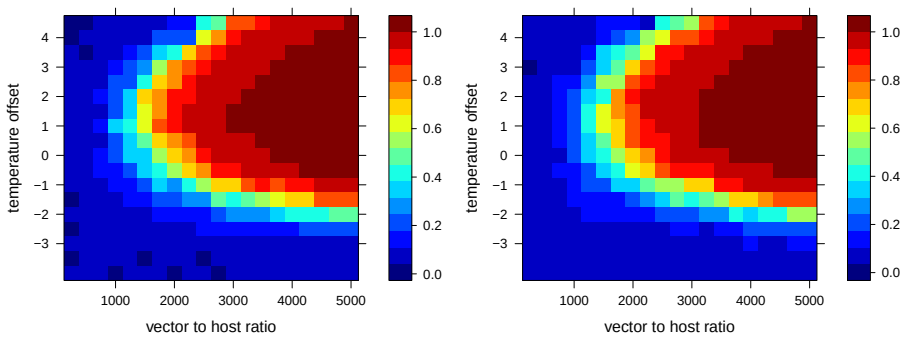

Figure 3.11: Final seroprevalence as a function of vector to host ratio and temperature offset. Temperatures are from Denmark 2008 (left) and Holland 2007 (right), with start day 150. The value presented is the median value of 100 simulations. Temperatures in NL and DK following day 150 are very similar, therefore the figures in this plot are also very similar.

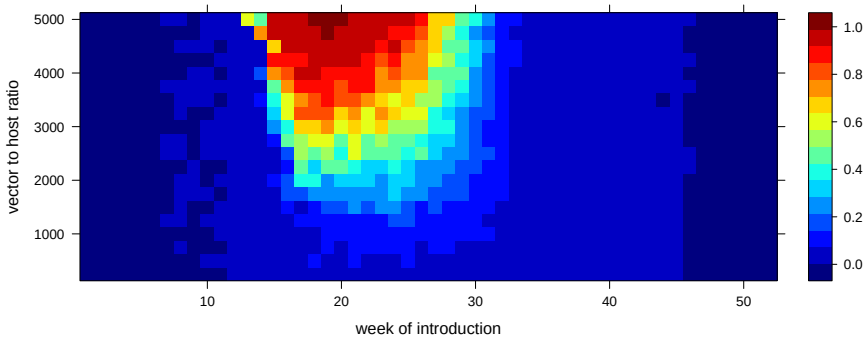

Figure 3.12: Final median seroprevalence as a function of week of introduction and vector to host ratio for DK temperatures. Figure 3.9 is produced in parts by the same data as presented in this figure. Each colored square represents the median value of 70 runs using the only the within-herd model. High values of seroprevalence in herds following an introduction of infectious midges can only happen when introduction happens in spring or summer.

Chapter 4

Conclusion & Discussion

The outcome of this project is a generic model which is able to simulate non-contagious infectious vector-borne disease between up to two species of hosts spatially distributed on farms and pastures. The model features the most realistic within-herd model of disease, combined with a triple movement kernel that describes spread of disease using midges or hosts as agents of the spread. There exist spread models of greater complexity, but none of these incorporates the same level of details of within-herd spread, they are i.e. not fully temperature dependent. When midge flying parameters have been quantified by experiments, this model can be implemented on areas naïve to the modeled disease with a high predictive power.

Furthermore this PhD has provided a new method of estimating the effect of light traps, which can estimate the additive effect of closely placed traps. There has also been devised a method to sample in time which maximizes information about time dependence and is robust to changes.

It was demonstrated how an epidemic outbreak of bluetongue in Denmark is sensitive to the use of pasture, climate, vaccination, vector abundance, and flying parameters. In this section these results are discussed compared to other models, and in relation to other research in this field.

The major finding from paper A is that an epidemic will be smaller when hosts

are placed in stables. This was interpreted as an effect of vectors having to fly longer to locate herds, and a longer time 'in transit' means longer time were they are in risk of dying. Stabling cattle could therefore be an effective means of preventing epidemics in Denmark. This model does not include an eventual added effect that housing may have on an epidemic by changing the biting frequency of midges. Some studies have indicated that vectors are not found in stables in as high numbers as on surrounding fields [74], while other indicates the opposite [69,75], which may dependent on *Culicoides* species [76]. If stabling further reduce the access of vectors to hosts, then the effect of stabling will be greater than predicted from the model. Stabling of animals could also influence the midge abundance, e.g. how manure is handled may affect breeding habitats.

In section 3.2.3 it was discussed that the present climate in Denmark is not sufficient to explain the small epidemic (15 cases) in 2008, but it was also showed that data indicates that increasing average temperatures will increase the size of epidemics. It is important to clarify that these findings only regard the temperature dependent parameters defined in the literature, which is the biting rate, the survival rate of midges, and the extrinsic incubation rate. How climate affect the abundance of midges may be the most important factor and this is also the parameter that is described in least detail. The National Veterinary Institute is currently building a model of midge abundance that includes many factors which should improve our understanding of midge abundance. These future results will be very interesting to include in the model. It is difficult to conclude whether the low number of bluetongue infected cattle in Denmark is due to these factors or the late introduction of the disease. Climate changes may however influence the temperature driven parameters. In paper A the sensitivity analysis (figure A.3) reveals that the size of an epidemic will increase with a factor of five if the temperature rises with one degree.

Given that the Netherlands and Denmark differ little in temperatures, and if it is assumed that time of introduction is the same, and that level of vectors to hosts are the same there is still one factor that is significantly different between DK and NL: the density of host animals are much higher in the Netherlands. According to the agricultural census in DK/NL 2007 (epp.eurostat.ec.europa.eu) NL has approximately nine times more sheep and more than twice the number of cattle compared to DK, in a country of the same size. It is seen in figure 3.2 that an introduction in a host dense area will lead to a larger outbreak. Furthermore the large amount of sheep in NL is most likely grazing, and also NL has a higher percentage of beef cattle, which is more likely to be on pasture. A higher percentage of grazing animals can also result in a larger epidemic (as concluded in paper A).

In general it can be difficult to obtain exact pasture distribution even in Denmark where each cow is individually registered and every movement must be

tracked, since no one registers whether cows are stabled or put to pasture. There are some indicators given the type of farm and the intended use of the cattle. Cattle for meat production are often placed outdoors, while dairy cattle are most often placed indoors, except that organic dairy cattle must have access to pasture part of the year. Considerations along these lines can help to estimate sampling procedures when modeling hosts on pasture in the model.

The new flying parameters included in this model are at the same time the strength and weakness of the model. It is the inherent strength because how midges fly determines how virus spread, and once these parameters are well determined then the predictive power and the generality of the model will be high. Presently however many of these flying parameters are not well determined and therefore it is difficult to make predictions without large margins of error. It was expected that many of these parameters would be determined by a parallel PhD project as mentioned in section 2.4. However the parameter estimation project was unfortunately slowed by a huge workload on data collection.

In the results section (3.1.1) it was observed that the flying parameters could result in under- or over-dispersion. The spread of virus is primarily dependent on vector ability to detect hosts, which is represented in the flying parameters. Hence the parameter values that gives the largest epidemic are also the parameters that indicate when midges are best at locating hosts. Therefore an evolutionary pressure towards these parameter values is likely. The parameter values used when running the simulation is typical close to the maximizing values.

That the output of the model depends on the abundance of vectors (as demonstrated e.g. in figures A.3 & A.4) seems trivial given that vectors are the agents of the disease. This is, however, not always the case for transmission kernel models: as we observe in paper B if the transmission between farms is not depending on the vector abundance, then the final size of the epidemic will neither. The reason is that a simple transmission model only operates with infectious and non-infectious farms, so as long as there is enough vectors to keep the within-herd R_0 larger than one, then the spread of disease is independent of whether there is one or one hundred infectious hosts on a farm. For this reason the model presented by Turner *et al.* [50], which have transmission probability proportional to number of infectious hosts, is better than the model by Gubbins *et al.* and Szmaragd *et al.* [45, 46] to predict the spread. However Turner *et al.* [50] does not model the within-herd prevalence which may be an oversimplification, as seen in section 3.2.4 within-herd prevalence is dependent on midge abundance. If it is then assumed that in almost all cases an incursion of bluetongue leads to a 100% infected farm, which is probably an overestimation, it will also affect greatly any conclusion made about host movements, as the impact of moving animals between farms will certainly be influenced by whether

the disease reaches 100% or 10% prevalence on each farm.

The benefits of using movement kernels compared to transmission kernels have been discussed throughout this thesis, and it was demonstrated in section 3.1.1 and in paper B how the model is more sensitive to midge abundance and vaccination, also the model shows less variability and is more stable to initialisation procedures compared to the transmission kernel model by Gubbins *et al.* [45]. As discussed in the previous paragraph some of these shortcomings are addressed by Turner *et al.* [50], however the general problem with not using vectors as agents to spread the virus is that for each transmission event virus instantly appear some distance away from the origin. This happens without a corresponding loss of virus from the origin, moreover it does not reflect the loss of virus that takes place between the transmission sites. To the effect that the amount of virus is overestimated both at the sending and the receiving farm. This overestimation of virus presence can possible then reduce the transmission probability in the kernel, which might explain the difficulty of initialisation in the model by Gubbins *et al.* [45].

In general transmission kernels can be good for modeling repeated introduction of a disease in the same area, given that all information about the spread of a disease is encompassed in the kernel. However it is very difficult to estimate the effect of changing spread parameters when using a transmission kernel. Spread parameters which could change from year to year include: Temperature, host density and distribution, midge abundance, host immunity (vaccination or natural), or change in policy towards a disease. Should changes in any of these parameters occur, the previous fitted transmission kernel may reveal little on future spread. In comparison process oriented models are good even if variables are changed within the same area, given that an agent based approach may better describe this change, because processes involved in spread are modeled individually and can be changed accordingly. Moreover process oriented models are good for introduction in new areas, because we have no prior information about the transmission pattern in the area and should therefore use only general information about temperature, vector behavior, host distribution, wind patterns, etc. This should provide a much better estimate than using transmission kernels fitted to other areas.

In section 3.2.4 it was demonstrated how temperature and vector to host ratio can determine the in-herd prevalence of disease. Within-herd modeling is not very computationally demanding and the parameters are reasonably well determined, therefore it should be incorporated in most process based simulations of disease.

In section 3.2.1 it was noted that the parameter values given for Schmallenberg in the EFSA rapport [36] should give Schmallenberg a less spread than BTV,

because the viraemic period for hosts are one third compared to BTV, however farther spread of SBV is observed. Therefore some parameters not yet reported must be in favor of SBV, either that or the restrictions put in place during the bluetongue epidemic were enough to limit the spread. The parameters in this model that can give a larger spread of disease using the same vectors are primarily: the transmission rate from host to vector, and possibly the extrinsic incubation rate. The first parameter can be a factor of ten higher compared to the value used throughout this thesis. The second parameter is beyond my expertise to estimate. It could also be that SBV transmit directly between hosts.

The models and methods which have been developed during this PhD project all have the potential to be used in future work within epidemiology (the sampling model proposed in paper C even more generally). In the next chapter some suggestions for immediate use are presented.

Chapter 5

Perspectives

Process oriented agent based models are becoming more and more frequent in many fields (personal experience from physics, epidemiology, and statistics, 3M+ Google scholar hits, and numerous articles that tries to define a standard for agent based models which I will not claim to have read). It is to be expected that this development continues given that computer power becomes cheaper and more accessible. The model presented in this thesis, the SWOFT [59], and the NAME [58] models (section 1.2 at p. 6) are examples of this development towards agent based stochastic modeling.

The model presented here is immediately applicable to other vector borne disease spreading in northern Europe, as was demonstrated in the case of Schmallenberg virus. Other than testing the model with diseases such as Akabane, Epizootic Haemorrhagic Disease, Douglas, Shamonda, Sathuperi or any of the other cattle diseases or African Horse Sickness and Equine Encephalosis Virus affecting horses all spread by midges, there is a still unfulfilled potential in testing with BTV and SBV e.g. unusual vaccination scenarios. An example could be to test mosaic vaccination where every second farm is vaccinated to lengthen the distance between farms, which would be an added effect comparable to going from pasture to farms. Another possible vaccination scheme could be to only vaccinate cattle on pasture, or if the Danish midge model proposes hot spots of midges to vaccinate cattle around these hot spots. The model could be used to explore these possibilities without modifications. It could also be

assumed that insect repellant acts as a simple factor reducing biting rate, while insecticides lower the abundance of midges. Any combination of the previous mentioned scenarios could also be tested, and an economic module could be added to evaluate the monetary effects of intervention strategies.

An interesting expansion to the model could be to investigate which parameters need to be different in order to explain the extraordinary spread of the Schmallenberg virus. Is it the transmission probability that differs, is there horizontal transmission of disease, what is needed to explain the disease given that the vectors are the same but the spread pattern is different? Or is the spread patterns of BTV and SBV not significantly different? The spread pattern of SBV is perhaps actually within spread scenarios that are plausible also for BTV.

It seems also obvious to incorporate a more exact interpretation of long distance spread. I.e. combining the model with the NAME model [58] (section 1.2). However the first step along this path is to get the medium length scale steps more accurate by including actual wind data. It would also be interesting to obtain better information about the use of pasture, by e.g. sending questionnaires to farmers, or to identify hosts on pasture using Google Maps.

Furthermore a different use for a model of this kind could be an option to run it 'backwards' So that when future incursion occurs, the model can point out possible routes of introduction, and more importantly possibly unidentified infected farms. Running the model 'backwards' could be implemented in several different ways. One could be actual backtracking of wind patterns and cattle movements that leads to the observed disease pattern. Another method could be to 'brute force' simulate all likely introduction scenarios to predict the most likely. Running a model 'backwards' first might also improve on forward prediction afterward, as the 'backward' run may reveal a number of non-detected virus spreaders which could influence forward spread severely.

Further improvement of the estimation of disease spread in Denmark could be done by including the results from "Opbygning af dansk veterinært beredskab for bluetongue" (Improving Danish veterinary preparedness against bluetongue), which is a cross institutional collaboration where one of the work packages has the purpose of determine factors that influence the abundance of midges in Denmark. This work package is due to finish within the year 2012 with a complete temporal and spatial model of midge abundance in Denmark. The results show that midge abundance is one of the key factors of an epidemic (paper A), and having better estimates of this parameter increases the predictive power of the model.

It would be of great delight to me if the model/code presented in this PhD thesis could be of use here in Denmark or elsewhere. As mentioned in appendix

E the code can be obtained in part from the net or in full from the author, as can the specifications for input data. The code is written in Fortran 90, which means it could be (without too much pain) converted to C or R with reasonable effort, and probably also with no loss of speed. It would also be possible to generalize the model further to take a higher unspecified number of vectors and hosts. Such a restructuring of the code may also make possible at speed boost by implementing more array operations. The possibility is there for someone to take...

Appendix A

Simulating spread of Bluetongue Virus by flying vectors between hosts on pasture

---

Kaare Græsbøll, René Bødker, Claes Enøe, Lasse E. Christiansen.

23 May 2012: Submitted to *Scientific Reports*.

13 August 2012: Received comments from reviewers.

31 August 2012: Submitted with changes according to review.

2 November 2012: Accepted for publication.

16 November 2012: Published.

Abstract

Bluetongue is a disease of ruminants which reached Denmark in 2007. We present a process-based stochastic simulation model of vector-borne diseases, where host animals are not confined to a central geographic farm coordinate, but can be distributed onto pasture areas. Furthermore vectors fly freely and display search behavior to locate areas with hosts. We also include wind spread of vectors, host movements, and vector seasonality. Results show that temperature and seasonality of vectors determines the period in which an incursion of Bluetongue may lead to epidemic spread in Denmark. Within this period of risk the number of infected hosts is affected by temperature, vector abundance, vector behavior, vectors' ability to locate hosts, and use of pasture. These results indicate that restricted grazing during outbreaks can reduce the number of infected hosts and the size of the affected area. The model can be implemented on other vector-borne diseases of grazing animals.

Introduction

Culicoides-borne diseases represent a growing expense for the European dairy and meat industry [35, 77, 78]. The outbreak of Bluetongue in the temperate parts of northern Europe in 2006-2008 is one of the costliest examples hereof. Recently *Culicoides* species from temperate parts of Europe have been shown to carry the Schmallenberg Virus [34], which is closely related to Akabane [33]; similarly to BTV, these two diseases have negative effects on ruminant production animals. Given the increased numbers of diseases carried by *Culicoides* it becomes important to develop strong decision-making tools to help assist in evaluating best practises against epidemic spread of vector-borne diseases. Such tools should be based on models that are able to predict the spread of disease with good precision.

Bluetongue virus (BTV) first occurred in Denmark in 2007. The arrival of the virus gave rise to only a few cases of BTV infection in that and the following year, however, a costly vaccination campaign was launched to prevent a wide-scale epidemic. Due to the relative shortness of the warm period in Denmark and the other Nordic countries, it has been hypothesized that the temporal window for BTV transmission may be so short in these countries that a large-

scale epidemic is unlikely, and hence vaccination or other preventive measures may not be necessary.

The aim of the model presented in this paper is to describe the spread of vector-borne diseases with a high degree of spatial precision. The model should also be so fast that it can be used to examine several different scenarios should such a vector-borne disease enter the area of interest, and simulate this within a reasonable amount of (CPU) time.

Following the 2006 outbreak of BTV in northern Europe, models with increasing complexity have been created to model Bluetongue [4, 45, 46, 50, 59, 79]. We elaborate on existing frameworks by introducing more realistic spatial parameters. Firstly, we assign host animals to pasture areas, and secondly, we simulate vector movements. By modeling the actual movements of vectors between pasture areas with and without hosts, we observe that spatial spread becomes more sensitive to parameters affecting vectors.

In the following we present more than 1,000,000 simulated years of Bluetongue in Denmark; the total data set represents less than 20 hours single CPU time.

Results

A major difference between our model and previous models is that in our model virus is transported between herds on pasture by vectors as opposed to between farm locations only. Therefore the area in which vectors can locate hosts with certainty becomes a large driving factor of epidemics. The simulations show that, when increasing the grid cell size, which represents the area in which vectors locate host animals, the number of affected cattle increases, as seen on rows in figure A.1. In the literature there are no direct observations of how large the area in which vectors will certainly locate hosts is. Based on mosquito data, Sedda *et al* [59] estimated that midges can detect hosts that are up to 300 meters away. Another way of increasing the probability with which vectors locate hosts is to increase the proportion of farms with host animals on fields. This increases the overall area with hosts and therefore also the probability for vectors to locate hosts. This can be observed in the increased number of affected cattle when there is increased use of pasture (columns in figure A.1). However, when the capability of vectors to locate hosts (the grid cell size) exceeds the average distance between neighboring farms, there will be no added effect of putting host to pasture because vectors can locate from farm to farm as easily as from pasture to pasture.

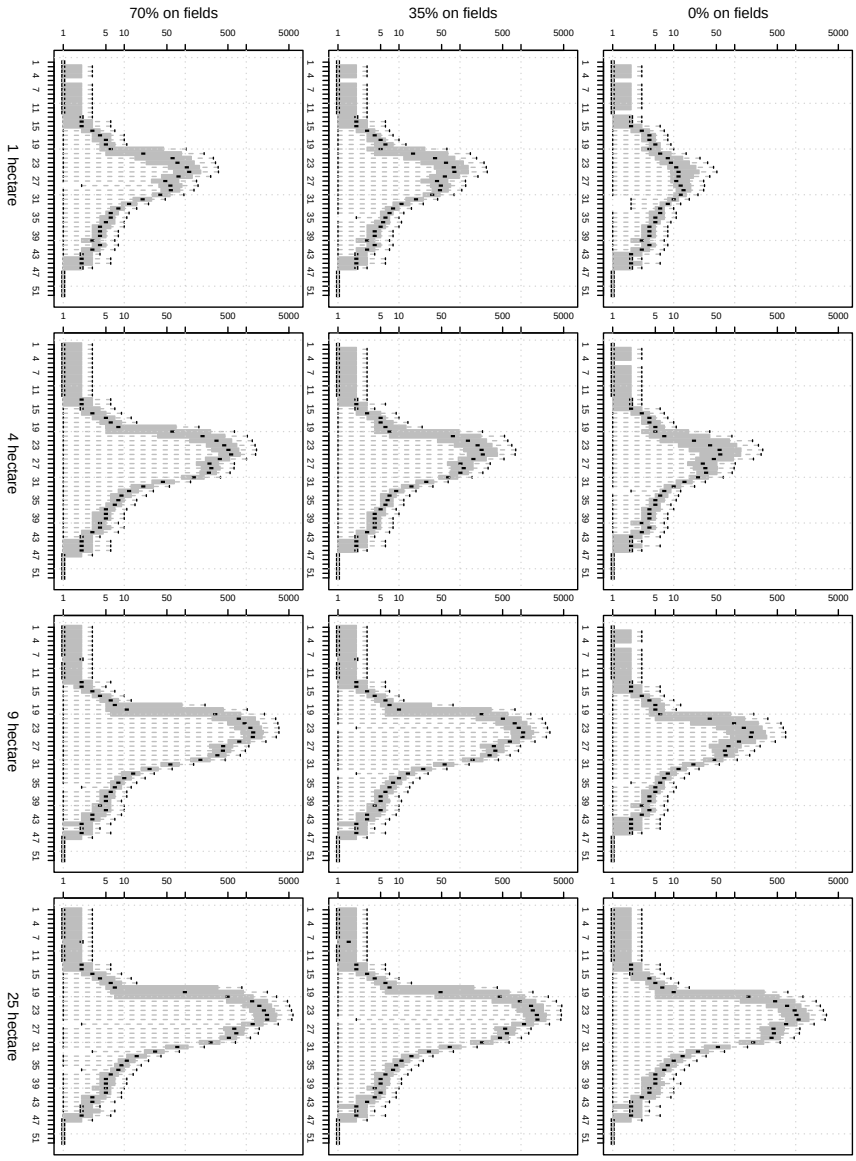

Figure A.1: Number of affected cattle as a function of week of introduction of disease. Rows have cattle distributed from 0% of cattle grazing on fields (all within farm buildings) to 70% on fields. Columns have the area in which vectors are able to locate hosts ranging from 1 to 25 hectare. One is added to all data to use log-scale on the y-axis.

The number of affected animals generally correlates positively with the spread of the disease (compare figures A.1 and A.2). Notice that the colored pixels in figures A.2 and A.4 are aggregated data 5 by 5 km and does not represent the grid cell size chosen in the simulation. The beginning of the epidemic outbreak window coincides with the beginning of the vector active period in week 20. Arrival of infectious vectors later than week 30 results in few disease incidences because the temperature is too low for the virus to complete its extrinsic incubation period in the vector. The combined effect of the short vector period and cold winters in Denmark results in a window of three months in which at least 50% of simulations show more than 500 cattle are affected with Bluetongue, given that at least 35% cattle are on pasture and vectors able to locate cattle on at least 9 hectares (figure A.1). The total number of affected hosts is very sensitive to additional parameters which we will describe in the next paragraph.

Sensitivity of the model. We tested the sensitivity of several parameters. In figures A.3 and A.4 (**a,b**) the number of infected host animals and the spread of disease is seen to be very dependent on the temperature and the number of midges per host. Higher vector abundance introduces more susceptible bites to infectious hosts, and higher temperature increases the probability for the virus to complete the EIP in the vector. Therefore, both higher vector abundance and higher temperature result in an increase in the number of infectious vectors. The increase in the total number of affected hosts increases with almost one order of magnitude every time the abundance of vectors doubles or the temperature rises one degree.

For the high temperature and high vector abundance scenarios, we observe higher presence of disease in coastal regions, especially along the edge of the simulation box. Part of this increase is an effect of boundary conditions in the model. If a vector is sampled to move outside the simulation box, this move will be canceled, and therefore vectors close to the edges have lower probability of moving. The edges of the simulation box are the same as the edges of figures A.2 and A.4. Furthermore, as it is not fatal for vectors in the simulation to be in areas above water, we expect many of the vectors that blow out to sea will return, thereby also increasing the presence of disease in coastal areas.

Figure A.4 (**c,d**) indicates that longer flight ranges, whether active or windborne, increase the spatial spread of disease. However, this increased geographic spread is not reflected in the number of infected animals. From figure A.3 (**c,d**) we observe that there are peak values in the number of affected hosts as a function of flying parameters. Beyond these peak values the simulations show fewer infected host animals. We therefore conclude that high dispersion of vectors may have a thinning effect on an epidemic outbreak (over-dispersion). This thinning effect, where the virus is spread over a greater area, but affects

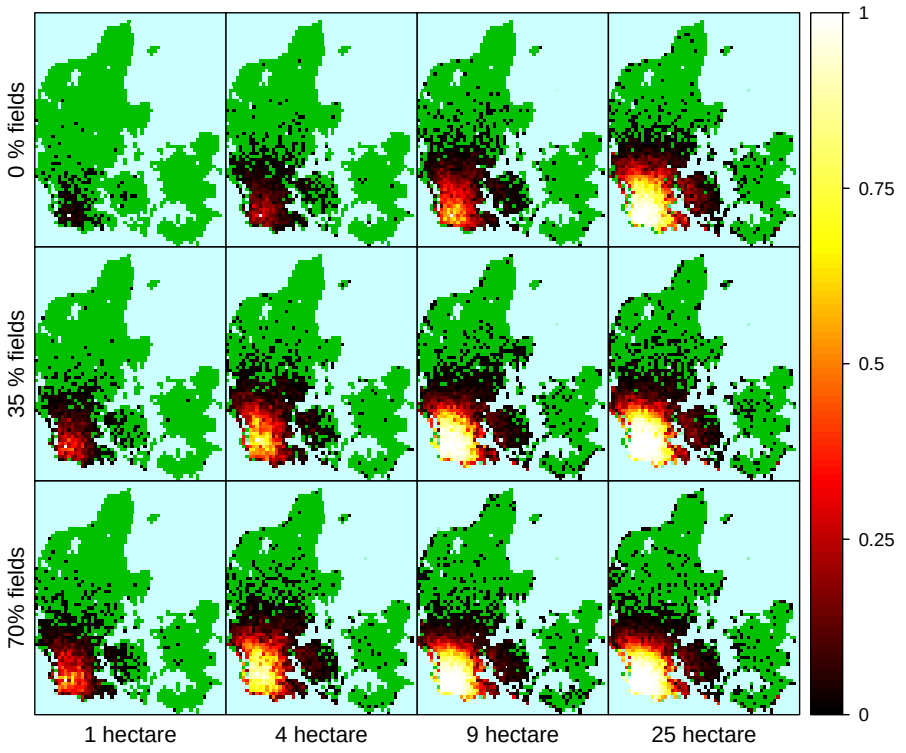

Figure A.2: Influence of vector homing range and cattle distribution on spread of Bluetongue Virus. Fraction of simulations that show presence of disease within a 5 by 5 km pixel for introduction of infection at day 170 repeated 100 times. Rows have cattle distributed from 0% of cattle grazing on fields (all within farm buildings) to 70% on fields. Columns have the area in which vectors are able to locate hosts ranging from 1 to 25 hectare. The color green indicates pasture were no cattle were infected with Bluetongue. The color blue indicates area with no Danish pastures.

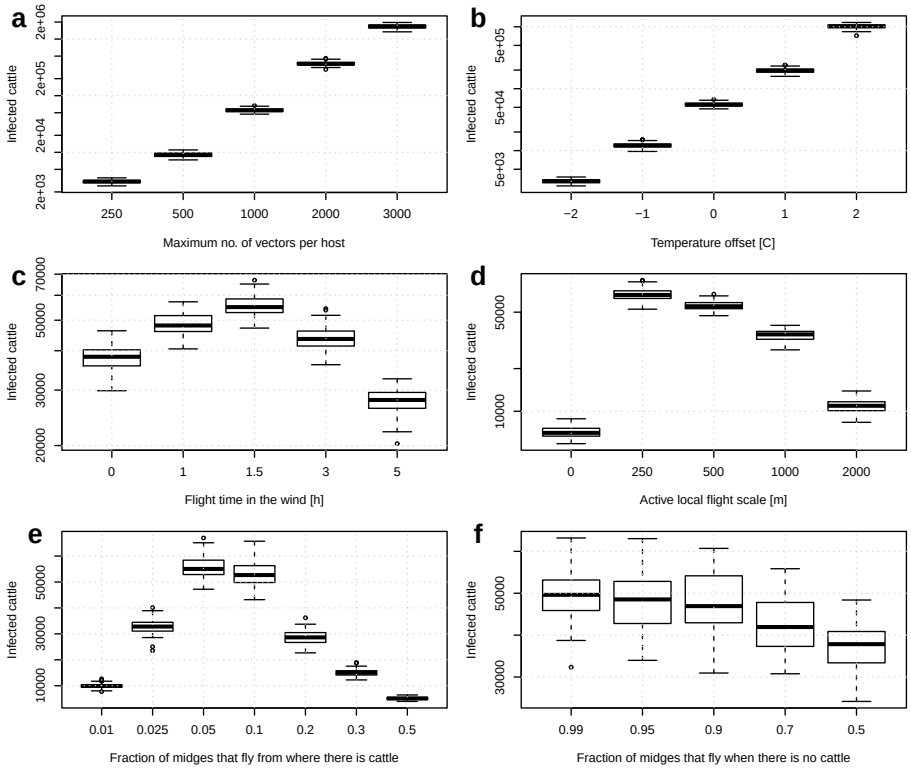

Figure A.3: Sensitivity of simulations. The total number of infected cattle for infection day 1 through 365 repeated 100 times as a function of: (**a**) the maximum number of vectors to host, m_M; (**b**) the 2008 Danish temperatures with offset [°C]; (**c**) the flight time in the wind [h], t_W; (**d**) the active local flight length [m], d_{L50}; (**e**) the fraction of vectors leaving area with hosts [1/day], η_H; and (**f**) the fraction of vectors leaving area with no hosts [1/day], $\eta_{!H}$. Grid cell size were 300 by 300 m, with 35% of hosts on pasture.

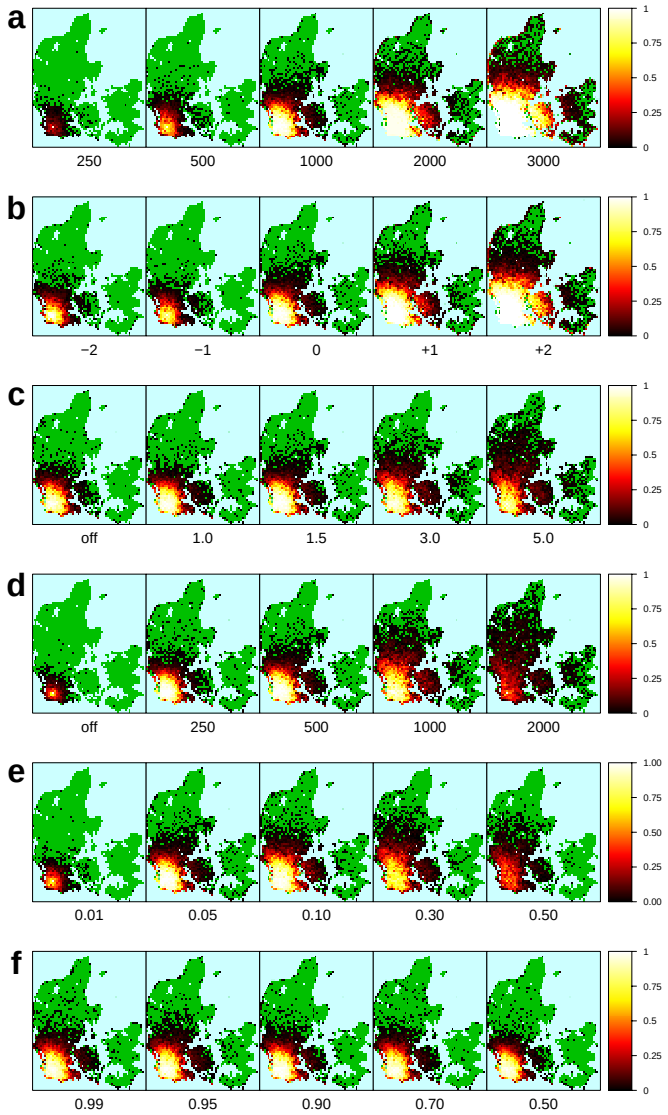

Figure A.4: Spread of Bluetongue virus. Fraction of 5 by 5 km pixels that show presence of virus for infection day 170 repeated 100 times as a function of: (**a**) the maximum number of vectors to host, m_M; (**b**) the 2008 Danish temperatures with offset [°C]; (**c**) the flight time in the wind [h], t_W; (**d**) the active local flight length [m], d_{L50}; (**e**) the fraction of vectors leaving area with hosts [1/day], η_H; and (**f**) the fraction of vectors leaving area with no hosts [1/day], $\eta_{!H}$. Grid cell size were 300 by 300 m, with 35% of hosts on pasture. The color green indicates pasture were no cattle were infected with Bluetongue. The color blue indicates areas with no Danish pastures.

fewer hosts can be explained as follows: the daily probability of infecting a host is $\lambda = I_V ab/H$ (from table A.1). For two herds with the same number of hosts, $H = H_1 = H_2$, and susceptible hosts, $S_H = S_{H1} = S_{H2}$, the probability that two infectious vectors present in one location, $I_{V1} = 2$, $I_{V2} = 0$, will not result in an infectious host is smaller than the probability that two infectious vectors in two different locations, $I_{V1} = I_{V2} = 1$, will not result in an infected host, $(1 - 2ab/H_1)^{S_{H1}}(1 - 0)^{S_{H2}} < (1 - ab/H_1)^{S_{H1}}(1 - ab/H_2)^{S_{H2}}$. The shorter the vectors fly the more probable it is that they land in the same place, and therefore the lower the risk of no infection of hosts. However, if they do not fly away from the place they acquired virus, with time the number of susceptible hosts S_{H1} will decrease because some of the hosts will become infected, and the inequality will no longer hold.

The sensitivity of the simulations to the probabilities of leaving areas with and without host animals (η_H and $\eta_{!H}$) are displayed in figures A.3 and A.4 (**e,f**), where we observe that the number of affected host animals is much more sensitive to whether vectors remain in a location with host animals, than to whether they leave areas without host animals. When a vector moves from a grid cell, the probability of it locating hosts is low (this probability varies with the distribution of hosts and is also geographically very different across Denmark because of differences in land use). As a result, the total number of affected cattle in the simulation is not very sensitive to changing the fraction of midges that fly from areas where there are no cattle, because most of these moves will result in landing in another cell without hosts. It is an *a priori* necessity for midges to fly from cells with hosts to achieve between-herd spread. Therefore figure A.3**e** displays an increase in the number of affected hosts when increasing the fraction of vectors leaving cells with hosts. This upward trend peaks around $\eta_H = 0.05$ day^{-1} because a large proportion of the infectious midges that fly away will never find new hosts, and therefore the number of affected hosts decreases if a large proportion of infectious vectors fly away instead of staying in the same grid cell and contributing to in-herd spread of disease.

With the exception of $\eta_{!H}$, we observe that the simulation model is very sensitive to all of the tested parameters (figures A.1–A.4), either with regard to the total number of cases or with regard to the spatial spread of the disease.

We have not conducted sensitivity analysis on the duration of viremic period on cattle, since this parameter is well defined in literature. Furthermore, no sensitivity analysis was conducted on transmission probabilities (b and β). Given that these parameters influence the number of infectious bites in the exact same way as the number of vectors per host, we conclude that these parameters also influence outbreaks in the same way. Gubbins (2008) [45] performed uncertainty analysis on these in-herd parameters, which concluded that the uncertainties do not influence on the epidemic outbreak.

We tested the simulations with and without movement of infected host animals, but with the very low probability of movement, p_m, no substantial influence could be observed on the total number of affected host (results not shown). With host movement only a few long range spread events could be observed given that 86% of movements are within 30 km and Danish cattle is in general not moved a lot compared to e.g. British sheep [50].

It is generally observed that varying the parameters does not change the period in which introduction of virus would lead to an outbreak in Denmark. The exceptions to this observation are: that the end of the vector free period determines when onset of an epidemic is possible, and that the temperature determines when vectors can no longer become infectious. In this study we do not include the possible effects of rising temperatures on the vector abundance; warmer temperatures may lead to an earlier onset of vector season, with a wider window of opportunity for an epidemic.

Discussion

From this study we conclude that a large-scale epidemic is possible in Denmark given certain parameter values such as high temperature, high midge abundance, and ease with which vectors can locate host animals. But incursion of such an epidemic is limited by a time window which is dependent on vector period and temperature. The simulations show that incursions of BTV later than August cannot give rise to large epidemics. This is in agreement with the observed incursion of Bluetongue in Denmark in 2007, where one index case in October did not lead to any further spreading of the virus. Also an introduction in late August 2008 lead only to 15 known infected herds, with total 27 animals affected. However, the detected incursion in 2008 happened during the beginning of the Danish vaccination campaign against BTV [80], which makes this incidence inconclusive evidence against the possibility of an epidemic outbreak.

Our simulations cannot predict the scale of an epidemic outbreak of Bluetongue in Denmark because the parameters used to describe vector behavior are not very well-defined. Therefore, we cannot conclude on the best prevention strategy for Denmark. The simulations do indicate that a warmer year than 2008, or a warmer climate, will increase the size of epidemics. Furthermore, reducing the probability of vectors locating hosts reduces the total number of affected hosts in an outbreak. This can e.g. be achieved by moving host animals into stables should an outbreak occur. Future experiments will hopefully help us to better determine the flying parameters of midges, in order to make better predictions on epidemic sizes.

Despite these reservations this new simulation model has proven to be highly sensitive to all parameters, which is desirable if input of sufficient quality can be provided. Given that many parameters, such as the flying parameters, and both host and vector distributions, are often poorly described, we can only conclude on results that are robust across parameter values. But with the increased focus on biting midges as vectors more parameters should soon be available. Given that the model relies on data reported from the EU, this model can be extended to all of Europe as it can incorporate the differences in climate and herd distribution across the EU.

The novel approach taken here of simulating midges as agents of spread of the disease has lead to the spread of virus in this model being highly sensitive to parameters affecting midges. We observe that adjusting parameter values so that they produce more infectious midges causes more infected hosts and a wider spatial spread of disease. Furthermore this type of process based modeling allows to include actual wind data, which can better represent the spatial spread of epidemics given that many areas have predominant wind directions. However to improve the predictive value of our and similar models requires much more information about the flight behavior of vectors.

Methods

Modeling vector-borne non-contagious diseases has two primary tasks: To describe the in-herd spread and the between-herd spread. In this paper we define a herd as any group of suitable hosts for the disease which is spatially discernible from other groups of hosts. This definition results in farm holdings often being split into several herds when put to pasture.

Distribution of herds. Denmark was divided into grid cells, where the cell side length is an input parameter ranging from 100 to 500 meters. In each grid cell, a number of hosts and vectors can be assigned. The vectors can localize all the hosts within the same grid as itself, thus the grid size in the simulation is essentially a measure of the vectors' ability to locate host animals. The hosts within one grid cell belongs to one herd.

To distribute hosts on pasture in Denmark, we combining information from the Danish Central Husbandry Register (CHR) with data from the EU arable land subsidies program linking cattle farmers with potential pasture areas. From this data, 70% of cattle owners were identified as registered landowners of grass-covered areas. This information made it possible to model different 'host on pasture' scenarios ranging from 'all hosts staying inside farm buildings' up to

'70% of cattle farmers distributing hosts on pasture'. As we do not know exactly which farmers do and do not put their animals out to grass, we randomly sampled which farmers had put animals to pasture for every simulation. Cattle selected for pasture were distributed into the grid cells covering pasture owned by their respective farmers. All animals not put out to pasture were placed in the grid cell containing the coordinate of the owner's farm. The data set includes 21,877 farms with 1.6 million cattle, of which 16,364 owned 89,756 fields covering a total of 2.0 million hectare. We have not included sheep or goats in this simulation given that cattle outnumbers these by more than ten to one.

The simulations cover all of Denmark with the exception of the island of Bornholm, which was excluded due to its epidemiologically isolated location in the Baltic Sea.

Between-herd spread. There are three principal ways in which the disease can move between grids in the simulation: transportation of host animals, active flight by vectors, and passive (wind-borne) flight of vectors.

Transportation of animals was modeled using a distance kernel. Which states the daily probability of a herd having hosts moved, p_M, and specifies the distance [km] hosts were moved, d_M, should such a move occur. The parameters were extracted from the movement register in Denmark, which tracks all movements of cattle.

$$d_M(x) = 12 \ \log^2\left(\frac{1.01 - x}{1.2}\right) \tag{A.1}$$

Where x is drawn from a uniform distribution on $]0;1]$. Data on movements were from the period October 2006 to September 2007, and therefore this data is not affected by the movement restrictions implemented after the discovery of Bluetongue Virus in Denmark. Movement restrictions followed the EU guidelines with regard to BTV (ec.europa.eu).

Within each grid cell it is assumed that the vectors can locate the host animals with 100% certainty. Beyond the size scale of the grid, vectors are assumed to fly a random walk pattern to locate hosts. A random walk can be described by a Gaussian spread kernel, and the flight distance of a vector, d_L, can thus be drawn using the normal distribution in equation (A.2) as the frequency distribution, equation (A.3).

$$f_G(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{A.2}$$

$$f(d_L) = 2 \ f_G\left(d_L; 0, \frac{d_{L50}}{P_{50}}\right) \tag{A.3}$$

Where the constant $P_{50} = 0.675$ scales so that d_{L50} becomes the radial distance 50% of the vectors will fly within. The factor of two normalizes the distribution given that distances cannot be negative.

Similarly for wind-borne flight, d_W, the wind carries vectors, and the distance carried is drawn again using the Normal distribution as frequency distribution.

$$f(d_W) = 2\ f_G\left(d_L; 0, \frac{d_{W50}}{P_{50}}\right) \quad , \quad d_{W50} = t_W w_S \tag{A.4}$$

Where d_{W50} is the distance 50% of the vectors are carried within by the wind. This is a product of the wind speed, w_S, and the time vectors spend in the wind, t_W.

The direction of the active flight is drawn from a uniform distribution. While for wind dispersal the direction, θ_W, is distributed as a Gaussian centered on the wind direction, w_θ.

$$f(\theta_W) = w_\theta + f_G\left(\theta_W; 0, \frac{\theta_c}{P_{90}}\right) \tag{A.5}$$

Where $\theta_c = \pi/4$ and $P_{90} = 1.645$ are given so that 90% of the vectors are within a cone spanning 45° around the wind direction. The direction of the wind used in this paper was randomly sampled once daily, and the wind speed was sampled uniformly from 0-5 m/s. Future versions of the program will accept input of actual wind data. We do not expect actual wind data to influence much on the size of outbreaks. However, real wind data will probably influence the spatial extent and shape of the outbreak.

Both active and passive vector movement are applied on each individual vector in each grid cell.

Simulated vectors in this model are "free" to disperse in a series of jumps before transmitting the disease. As opposed to traditional kernel models where transmission events are determined by the kernel. Free dispersal in combination with splitting vector movements into two separate Gaussians makes it possible for transmission events to occur as observed in real epidemics both short, long and fat tails as dependent by regional farming practices or policies [47].

We introduce two parameters to describe the behavior of vectors in grid cells with and without host animals. The daily probability of vectors leaving cells with no host $\eta_{!H} = 0.95$ day^{-1}, is set at a high value to reflect that the evolutionary pressure to locate blood meals is very high. When a stable source of

blood meals is discovered, there is little gain in flying away, so the daily probability of leaving an area with host animals is set at a low value, $\eta_H = 0.05$ day^{-1}. When we assume that midges prefer to stay in grid cells with hosts, we implicitly assume that there are suitable breeding grounds within the grid cell or that vectors have a homing behavior, which enables them to migrate between the same herd and their breeding sites. The validity of these assumptions depends on the behavior of the vectors, and *Culicoides* are known to display different breeding patterns depending on species [81,82]. Other models without active movement of vectors generally have an implicit $\eta_H = 0$, because virus is not removed from herds in case of transmission events.

In-herd spread. Within each herd, the dynamics work in almost the same way as described in previous articles relating to Bluetongue modeling [45,46]. Hosts are described using an extended Susceptible Exposed Infectious Recovered (SEIR) type framework, while vectors are described using an extended Susceptible Exposed Infectious (SEI) framework (figure A.5). The extended framework ensures that the model emulates virus incubation time more realistically compared with the non-extended approach. The SEI/SEIR framework uses rates which assign a probability of movement of an individual vector/host from one state to the next within one time step. In the original unextended SEIR model ($k = 1$) most midges become infectious on the time step immediately after obtaining the virus, which in the case of Bluetongue (and many other vector borne diseases) is impossible in nature, especially in areas with low temperature resulting in a long extrinsic incubation period (EIP). Therefore the exposed states of vectors are modeled using extra stages to better emulate the incubation period. These extra stages bring the frequency distribution of vectors completing the incubation period from an exponential distribution to a Erlang (Gamma) distribution [17]. When adding a risk of dying to the system, the frequency of vectors completing the incubation period becomes a phase-type distribution [60]. We also uses an extended approach on the infectious period of the host to better describe the viremic period.

Figure A.5 depicts the stages in the extended SEI/SEIR model. The arrows indicate an event described by parameters and transition probabilities in table A.1 and A.2. Bluetongue is a non-contagious disease and therefore transmission of virus is only possible through the bite of the vectors.

Vector abundance. The spread of vector borne diseases is highly sensitive to the abundance of vectors, but the vector abundance is often not well determined. For the Bluetongue vector there are numerous papers on catches of midges; where in the field one may catch thousands of midges one night and close to zero on the next. Therefore modeling abundance is often difficult and most be done over longer timescales to account for the large variability in daily trap catches [66,83,84].

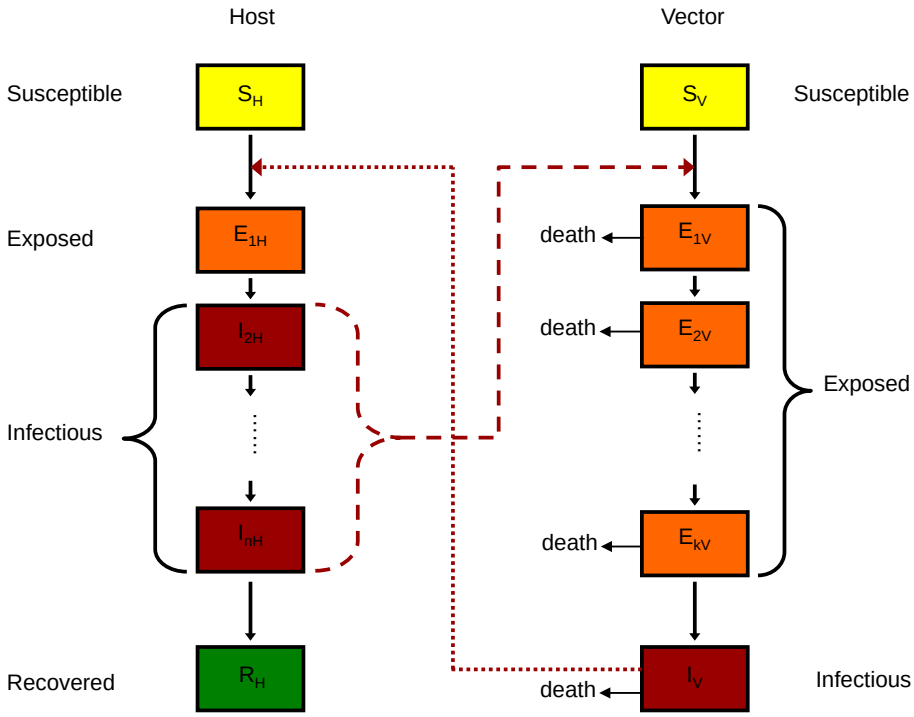

Figure A.5: Dynamic of disease. The viraemia of the hosts is described by an extended SEIR model, and the vectors are described by an extended SEI model. All movements between the stages in the model are governed by the probabilities listed in tables A.1-A.2.

The seasonal midge abundance, m, in Denmark across the year (table A.1) is adopted from Nielsen (1996) [65]. The abundance of vectors is represented by the absolute of a sinusoidal curve to emulate four generations from day 150 to 310 in the year. The estimate of m_M represents the maximum abundance of vectors which is multiplied onto the seasonality to give the daily maximum abundance. We sample the number of susceptible vectors uniformly from zero to daily maximum each day at each location to account for the day to day variation in midge catch that is observed in the original studies.

Simulations. All simulations were run for one year from 1 January to 31 December using temperature data from 39 selected meteorological stations recorded in 2008. Data from the 39 weather stations was interpolated to a grid of 25 by 25 km (71 grid points) in connection with another project (www.nordrisk.dk), and each herd used the temperature from nearest grid point. In the simulation

disease is introduced by 10 infectious midges that arrive in the southern part of Jutland (the large landmass in the western part of Denmark). The date of arrival is repeated 100 times across the whole year for each of the 12 scenarios for distribution of host and search area of vectors, in total $365 \cdot 100 \cdot 12 = 438,000$ years simulated for the data in figure A.1. The point of arrival can be seen as a white square in the plots in figure A.2. However, the exact spot of arrival can vary slightly between scenarios because we required the spot of arrival to have host animals, and presence of host animals depends on the distribution of hosts on pasture.

The introduction point in southern Jutland was chosen given that both Bluetongue (2009) and Schmallenberg virus (2012) have been previously introduced in this area (See the OIE WAHID database). Furthermore this area have the highest density of cattle in Denmark, and is therefore considered of highest risk. The climate in Denmark is very homogenous and introducing virus in other parts of Denmark gives similar results (not shown), although scaled with the regional host density.

Sensitivity analysis was done on the scenario with grid cell size 300 by 300 m and 35% of hosts on pasture. The tested outcome was the total number of affected hosts for a period of $365 \cdot 100 = 36,500$ years of simulated outbreaks per parameter value.

References

References are sorted numerically on page 145 and alphabetically on 157.

Acknowledgments

The authors greatly appreciate the Danish Knowledge Center for Agriculture, Dairy & Cattle farming, for providing farm and movement data from the Danish cattle data base and for providing part of the funding. The Danish Meteorological Institute for contributing to the project with temperature data from the national climate stations. And the Faculty of Agricultural Sciences at Aarhus University for providing pasture data. Permission to merge these data sets was granted by the Danish Data Protection Agency on 1 July 2008. This study was partially funded by the EU grant FP7-261504 EDENext and is catalogued by the EDENext Steering Committee as EDENext046 (http://www.edenext.eu).

The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

Author contributions

Conceived and designed the model: KG. Inputs on optimizing and structure of the model: LC. Inputs on biological behavior: RB. Wrote the paper: KG. Critically reviewed the paper: LC RB CE.

Additional information

Competing financial interests

The authors declare no competing financial interests.

Table A.1: Table of parameters.

symbol	value [range]	description [unit]	references
T	data	temperature in DK by 25km grid	—
H	data	herd size	—
a	$0.0002\,T\,(T-3.7)(41.9-T)^{0.37}$	biting rate [1/days]	[71]
b	0.9	probability of transmission from vector to host	[85]
β	0.1	probability of transmission from host to vector	[30, 45, 86]
m	$m_M\left\lvert\sin(\frac{t-150}{40}\pi)\sin(\frac{t-150}{160}\pi)\right\rvert$*	Seasonality of vectors	[65]
m_M	1000	maximum no. of vectors per host	—
$1/r$	20.6	mean duration of host viraemia [days]	[73]
n	5	stages in viremic period for hosts	[73]
$1/\nu$	$\nu(T) = 0.018\,(T-13.4)$ [1–100]	mean extrinsic incubation period (EIP) [days]	[17]
k	$0.5/\nu$ [3–11]	stages to describe EIP in vector	[17]
μ	$0.009\exp(0.160\,T)$ [0.05–0.9]	vector mortality rate	[72]
η_H	0.05	probability of vector leaving area with hosts	—
η_{IH}	0.95	probability of vector leaving area with no hosts	—
t_W	1.5	time vectors stay in the wind [hours]	—
d_{L50}	0.5	median length vectors will fly at random [km]	—
p_M	0.001	daily probability of hosts movement [1/herd]	—

Expressions dependent on temperature are imposed [range] limits shown in the table.
*For $150 \leq t \leq 310$ 0 for t otherwise.

Table A.2: Transition probabilities between states in figure A.5

probability of	expression	transition
infecting host	$ab\, I_V/H$	$S_H \to E_{1H}$
host becoming infectious	nr	$E_{1H} \to I_{2H}$
host moving through stages	nr	$I_{(i-1)H} \to I_{iH}\ i = 3, 4, ..., n$
host recovering	nr	$I_{nH} \to R_H$
infecting vector	$a\beta \sum_2^n I_{iH}/H$	$S_V \to E_{1V}$
vector moving through stages	$k\nu$	$E_{(i-1)V} \to E_{iV}\ i = 2, 3, ..., k$
vector becoming infectious	$k\nu$	$E_{nV} \to I_V$

Appendix B

A comparison of dynamics in two models for the spread of a vector borne disease

Kaare Græsbøll, Thomas Sumner, Claes Enøe, Lasse E. Christiansen, Simon Gubbins.

27 October 2012: Submitted to *Scientific Reports*.

25 January 2013: Submitted with changes according to review.

11 April 2013: Submitted with changes according to further review.

Abstract

In 2007 bluetongue was introduced to both Denmark (DK) and the United Kingdom (UK). For this reason simulation models were built to predict scenarios for future incursions. The DK and UK models share within-herd dynamics, but differ greatly in their descriptions of between-herd spread, one using an explicit representation of vector dispersal, the other a transmission kernel. Here we compare model predictions for the dynamics of bluetongue in the UK, based on the 2007 incursion and vaccination rollout in 2008. We demonstrate how an agent-based model shows greater sensitivity to the level of vaccine uptake, and has lower variability compared to a kernel-based model. However, a model using a transmission kernel runs well with less detailed data and is often faster. The two different models give similar outputs regarding the effects of vaccination, and to some extent the spatial spread of disease. These results strengthen the reliability of both models.

Introduction

The use of modeling is becoming an ever more important part of decision making for the control of infectious diseases in production animals across the world. Specific actions have been taken on the basis of models in several cases including foot-and-mouth disease [87], bluetongue [88], and avian influenza [89]. The objective of modeling is to predict the outcome of circumstances not previously observed and thereby inform the best action taken should such circumstances present themselves. Simulating highly complex systems, such as vector-borne diseases, the input data comes with large uncertainties especially in the locations and movement of vectors and to a lesser degree for the hosts. The level of realistic representation in a model should represent the level of knowledge we currently possess. Testing the validity of models is an obvious concern, and often poses some difficulties, especially if outbreaks are infrequent and the data available is sparse. Therefore, comparing models can act as an extra step of validation [87]. This is even more the case if models use different, but equally valid assumptions, so that concurring results may be interpreted as more robust.

In this paper we focus on models for the spread of bluetongue, a vector-borne disease. Statistical and mathematical modeling of vector-borne diseases, generally,

and bluetongue, specifically, is a wide field, including those for long range spread of vectors via wind [51, 53–55, 58], influence of climate and landscape [2, 48, 79], within-herd models [4, 43, 45], between-farm spread [47], and transmission within and between farms [46, 50, 59, 90].

Here we compare two models [46, 90] which share the same vector-host dynamics within a farm, but differ in their representations of spread between farms. One model is based on a probability kernel to evaluate risk of spread from farm to farm based solely on the distance between farms (referred to as the kernel-model) [46]. The other assumes host animals to be distributed on both farms and pasture areas, and assumes passive (wind) and active (flight) vector movements act as the primary mechanism of disease spread (referred to as the fly-model) [90]. Here we use these models to describe the 2007 outbreak of bluetongue in the UK, and the effect of proposed vaccination schemes. In this case study hosts are cattle and sheep, and vectors are *Culicoides* biting midges.

Results

The probability of an outbreak taking off (defined as any spread from the farms infected initially) differed significantly between the kernel- and fly-models. In the kernel model only 2.3% of introductions resulted in outbreaks, whereas in the fly-model all introductions resulted in outbreaks. Furthermore, the probability of an outbreak did not differ significantly for different vector-to-host ratios (sampled from a uniform U(0,5000) or fixed at 2500 or 5000). Results in the remainder of this section are conditional on the outbreak taking-off.

The two models both show that increased vaccination uptake reduces both the total number of infected holdings and the spread of the outbreak ($P<0.001$) (Figures B.1-B.3). In addition, the total number of affected holdings is not significantly affected by whether the vector-to-host ratio is sampled from a uniform U(0,5000) distribution or is fixed at 2500 (i.e. the mean of the sampling distribution). The variances of the outcomes (the size of the simulation envelope) for the fly-model are significantly smaller than for the kernel-model ($P<0.001$).

At low levels of vaccination (none or 50%) the weekly incidence of newly affected holdings is higher for the fly-model compared with the kernel model, though the predictions for the kernel-model are considerably more variable (Figure B.2). After week 50, the incidence in the kernel-model reaches a plateau, or at most shows a modest decline. This contrasts with the fly-model which shows a marked decrease in incidence after week 55. For higher levels of vaccine uptake (80%, 95% and variable), however, the incidence is higher for the kernel model than

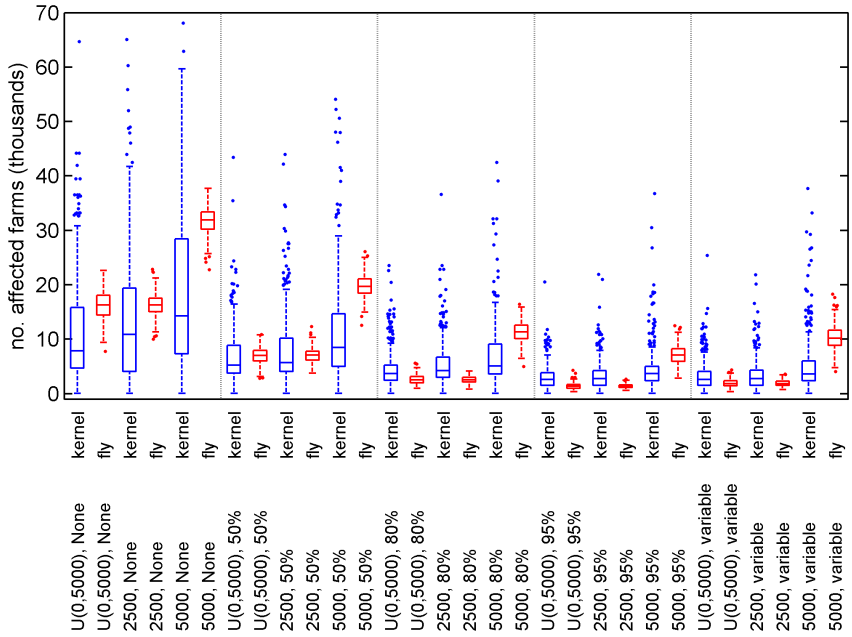

Figure B.1: Cumulative number of affected holdings and its dependence on vaccine uptake and the vector-to-host ratio. Scenarios are grouped by vaccine uptake: no vaccination, 50%, 80%, 95%, and variable (95% in the 2007 protection zone, 75% in the 2007 surveillance zone and 50% elsewhere); and by vector-to-host ratio: sampled from a Uniform(0,5000) distribution or fixed at 2500 or 5000. Box-and-whisker plots show the median (line), interquartile range (boxes), 1.5 times the interquartile range (whiskers) and any outliers (crosses). The kernel-model is indicated by blue boxes, the fly-model by red boxes.

for the fly-model (except when the vector-to-host ratio is fixed at 5000) (Figure B.2).

The predicted spread of bluetongue virus is very different in the two models (Figure B.3). The kernel model predicts a very concentrated presence in East Anglia regardless of vaccination uptake. However, vaccination substantially reduces spread from eastern England. By contrast the fly-model predicted a similar extent of spread for all vaccine scenarios, but the probability of spread within this region was substantially reduced by increasing vaccine uptake. Consequently, disease was predicted to be more widespread in the fly-model than the kernel model for high vaccine uptake, but since the number of affected hold-

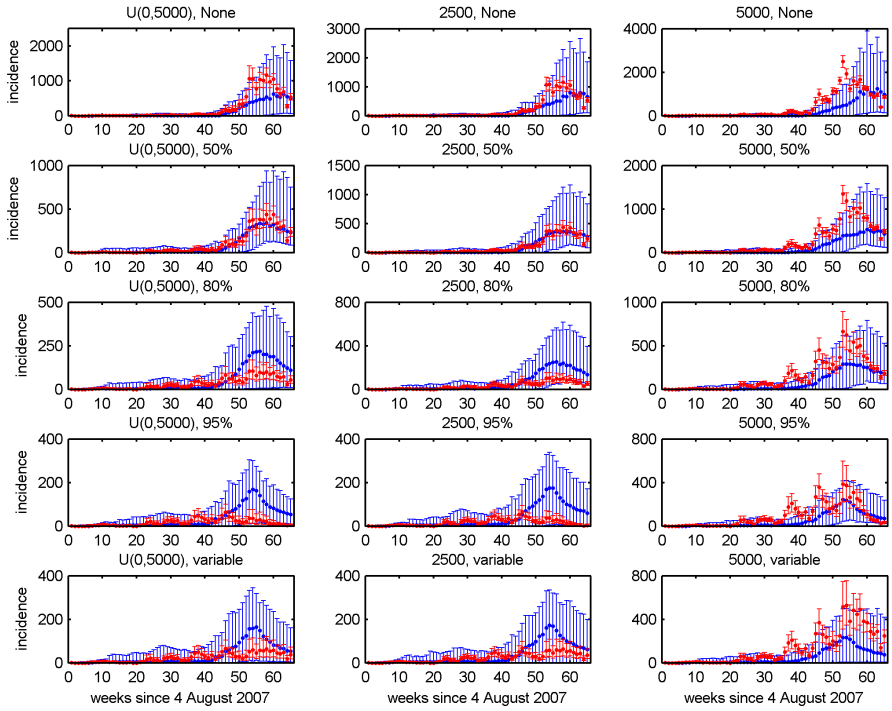

Figure B.2: Predicted incidence of newly infected holdings each week for the kernel-model (black symbols) and the the fly-model (red symbols). Each figure shows the median (circles) and 10th and 90th percentiles (error bars) for each model. Columns differ in the vector-to-host ratio: sampled from a Uniform(0,5000) distribution (left) or fixed at 2500 (middle) or 5000 (right). Rows differ in vaccine uptake (from the top): no vaccination, 50%, 80%, 95%, and variable (95% in the 2007 protection zone, 75% in the 2007 surveillance zone and 50% elsewhere).

ings (Figures B.1 & B.2) is lower in the fly-model, this implies that the virus has a lower holding prevalence in affected areas compared to the kernel-model (excluding the case where the vector-to-host ratio is 5000). Moreover, predicted spread in the fly model tends to be much patchier than in the kernel model, where transmission is much more uniform.

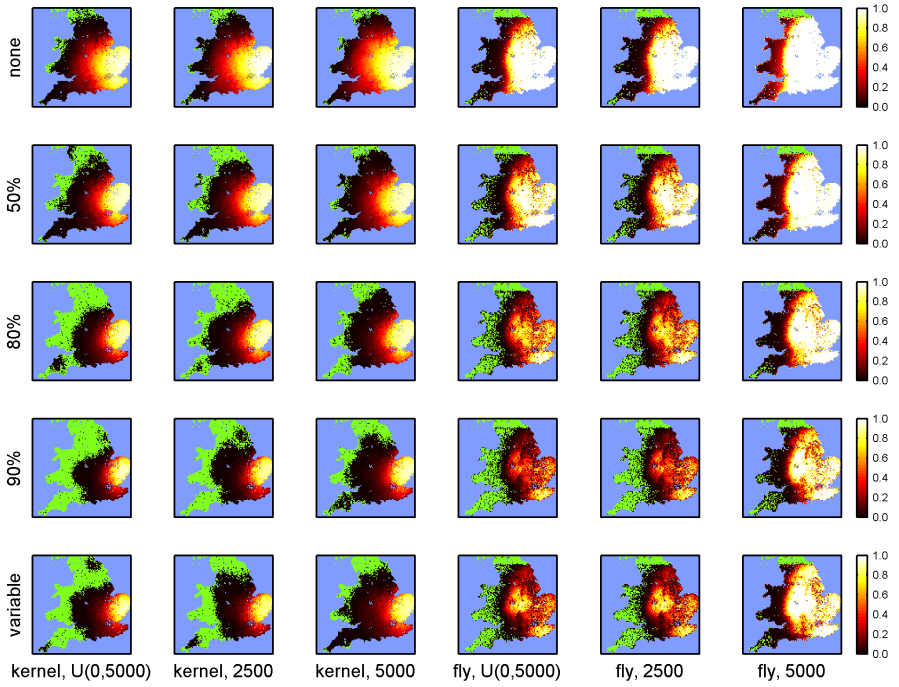

Figure B.3: Predicted spatial spread of BTV in Great Britain. Each figure show the proportion of simulations for which at least one farm became infected within a 5km grid square. The three left-most columns show results for the kernel-model, while the three right-most show those for the fly-model. Each column has the same vector-to-host ratio: sampled from a Uniform(0,5000) distribution or fixed at 2500 or 5000. Rows differ by vaccine uptake with no vaccination, 50%, 80%, 95%, and variable level of vaccine uptake (95% in the 2007 protection zone, 75% in the 2007 surveillance zone and 50% elsewhere). The fly-model was restricted from spreading to Scotland, so that the sharp northern cut-off in some plots are an artifact of this restriction.

Discussion

Testing different models on similar conditions can be an important step in model verification and validation. In this study we have compared the predictions of two models for the spread of bluetongue, the fly-model and the kernel-model, whose descriptions of spread between farms and of the distribution of livestock

are very different. The qualitative predictions of the two models in terms of the impact of vaccination are broadly comparable. This gives us some confidence that the conclusions made about the relative impact of control are robust and not strongly dependent on the assumptions made in the models. There are, however, some important quantitative differences between the models and understanding how these relate to the assumptions made in the model give insights into the transmission dynamics of bluetongue virus.

One noticeable difference between the models is that the predictions of the fly-model are more sensitive to assumptions about the vector-to-host ratio than the kernel model (Figures B.1-B.3). The fly-model depends on midges to transmit the virus between farms and, consequently, the size of an outbreak in this model is directly proportional to the number of vectors per host (Figure B.1). By contrast, spread in the kernel-model can occur as soon as one vector on a holding completes the extrinsic incubation period (EIP) after having acquired infection from a host. The probability of spread from farm to farm does not depend on the number of infectious vectors, so the number of affected holdings is not significantly different provided the vector-to-host ratio exceeds the threshold needed to produce the first infectious vector (i.e. that required for the basic reproduction number to exceed one; cf. Gubbins et al. 2008 [45]).

Because transmission between farms in the fly-model depends directly on the number of infectious vectors, this introduces temperature dependence to spread (because of the temperature dependence of the EIP), which is not the case for the kernel-model. This temperature dependence accounts for the decline in new cases later in the year predicted by the fly-model which does not occur in the kernel model, at least at low vaccine uptake (Figure B.2). Analysis of the 2006 outbreak in northern Europe suggested that transmission between farms required a daily mean temperature above 15°C [47], which is consistent with the temperature dependence of the EIP for bluetongue virus [17].

The patterns of spatial spread differ between the models (Figure B.3), which partly reflects their representation of the host populations and partly the way in which transmission is modeled. The patchy spread and low prevalence of affected herds in the fly-model are consequences of the distribution of hosts on pasture, which effectively means there is a larger number of smaller herds than for the kernel model. This results in a larger number of potential outbreaks each with a smaller probability of occurring. This will be further exacerbated by the dependence of transmission in the fly-model on the prevalence within a herd. By contrast, the kernel-model depends only on the presence of infectious vectors (not the number) resulting in much more uniform spread.

One further difference between the models is that the predictions of the kernel-model are considerably more variable than for the fly-model (Figures B.1 &

B.2). The kernel-model includes uncertainty in the parameters for the within-herd component, while the fly-model assumes fixed values, and two parameters (the vector-to-host ratio and the probability of transmission from host to vector) vary over several orders of magnitude. However, fixing either of these parameters at their mean value does not greatly reduce the variability in the model predictions (vector-to-host ratio: Figure B.2; probability of transmission, not shown). Consequently, the increased variability in the kernel-model is most likely a result of the description of between-farm spread, as this is the major difference between the two models. In particular, the distribution of hosts on to pasture in the fly-model increases the number and density of herds (pasture areas with hosts are being treated as herds). The distribution of hosts onto pasture allows the infection to spread in smaller steps, with vectors performing random walks between herds (which is similar to a continuous diffusion of virus), which reduces the variability in outcomes compared to the single step mechanism of the kernel-model.

The impact of the different descriptions of between-herd spread is also evident in the probability of an incursion taking off. For the kernel-model only 2.3% of introductions resulted in an outbreak (i.e. any spread from the initially-infected farms). This proportion is independent of the whether or not the vector-to-host ratio or probability of transmission from host to vector is sampled or fixed and is similar to that reported for the model previously [46,67]. By contrast, in the process oriented fly model introductions always result in an outbreak.

Using a transmission kernel has the advantage of simplicity and a kernel can be estimated directly from outbreak data without much difficulty. Although recent work has explored the potential for estimating the kernel as an outbreak develops [91], the kernel is typically derived from data on previous outbreaks, which raises the possibility that it may not reflect the current situation. This limitation can be addressed by investigating the sensitivity of the model predictions to the shape of the kernel used (see, for example, Szmaragd et al. 2010 [67]). A further issue is that a kernel combines all routes of transmission into a single mechanism, which can make it difficult to use a kernel-based model to predict the impact of control measures which influence only one mechanism (e.g. movement restrictions) [46,47].

Representing individual transmission mechanisms in greater detail (such as in the fly-model) has the advantage that it allows the roles of these mechanisms to be explored in more detail. In addition, it may be possible to estimate the relevant parameters prior to an outbreak, though this would require detailed information on the behavior of vectors which is not necessarily available. However, estimating the required parameters from outbreak data, while possible, does present a considerable technical challenge [59].

Recently, there have been a number of publications where the movements of midges have been modeled in greater detail. Turner et al. [50] modeled animal and vector movements, separately, Sedda et al. [59] evaluated short range flying patterns of midges in the 2006 BTV8 outbreak, and Burgin et al. [58] described the long range dispersal of midges by wind. These papers and the fly-model indicate a shift towards more process/agent based models and away from kernel-based approaches [46,47]. This is feasible for Europe were demographic and wind data is recorded at the level of detail needed. Hopefully, the results of such models will be more precise forward and backward predictions of transmission routes in future outbreaks of midge borne diseases.

Methods

The models used in this paper are presented fully in Szmaragd et al. 2009, 2010 [46, 67] (kernel-model), and Graesboell et al. 2012 [90] (fly-model). Here we provide a short resumé for each, with emphasis on the differences between the methods. We use the term 'herd' to describe any group of hosts spatially separated from other groups of hosts, thereby including host animals on farms and on fields. A 'holding' refers to all host animals registered under one County Parish Holding (CPH) number (the unique identifier used for holdings in GB), and can include multiple herds. A holding is attributed with the spatial coordinates of the corresponding farm. For this case study on bluetongue hosts are cattle and sheep; and vectors are any species of Culicoides biting midge capable of transmitting the virus.

Modeling framework

Both models are stochastic models that attribute a probability of an event (e.g. infection or death) happening and for each time step evaluate whether or not each event occurs in concordance with the Monte Carlo methods.

Spatial distribution of hosts: The kernel-model assumes point locations for each farm. The location and the number of cattle and sheep on each farm were extracted from June agricultural survey data for GB for 2006. The fly-model used the same farm data but also distributed host animals onto pasture. Pasture data were provided by the Rural Payments Agency (RPA), and combined with data from the 2006 agricultural survey data to provide a method of determining on which fields hosts animals could be grazing. However, only around 40% of farms were identified with fields through this process. All farms with identified

pasture had their host animals attributed to those. For farms not identified with pasture, all host animals were placed on the farm location (as in the kernel model).

Within-herd spread of disease was modeled by an expanded SIR-SEI construction by both models as presented in the original articles [46, 90]. Hosts were divided into Susceptible (i.e. uninfected), Infectious (i.e. capable of transmitting infection), and Recovered (i.e. no longer infected) states (SIR), while vectors were described by Susceptible, Exposed (i.e. infected but not yet infectious), and Infectious states (SEI). For a host to move from the susceptible to the infectious state it must be bitten by an infectious vector; after some time the host will recover from the disease or possibly die due to severe disease. Vectors move from the susceptible to the exposed state by biting an infectious host; the exposed state reflects the fact that the virus must replicate and move from the gut to the salivary glands of the vector. Virus replication is highly temperature dependent, and the timescale of this process therefore varies according to season. To best model the distribution of infectious time for the hosts and exposed time for the vectors additional I- and E-states were introduced for hosts and vectors to pass through before dying or reaching a recovered or infectious state. The implementation of these extended states differs slightly between the two models, but the authors verified that the outcomes in terms of waiting time distributions are equivalent. Another difference is that in the kernel-model parameters for transmission probabilities were sampled from intervals, whereas the fly-model used point estimates. Furthermore, only the kernel-model includes disease-associated mortality, but, because this is low, most animals survive the infection and those who do not die later in their infectious periods, having had opportunity to transmit infection to vectors. In the original paper [90] the fly-model used a SEIR model for the hosts; this has been simplified to match the SIR formulation used in the kernel-model. Because of the difference in host distribution the herd sizes are in most cases smaller in the fly-model. Given that the number of vectors is always proportional with the number of hosts, herd size does not influence the number of vectors infected by an infectious host and vice versa.

Between-herd spread of virus is the major difference between the fly-model and the kernel-model. In the kernel-model the daily probability of transmitting disease from an infected farm to an uninfected farm depends on the distance between the farms (with shape of this dependence described by a Gaussian function) and the demography of the farms (more precisely, whether or not they keep cattle or sheep) and used the parameters described in Szmaragd et al. 2009 [46]. The kernel-model implicitly includes all routes of transmission in one simple probabilistic description. When a farm becomes infected, the within-farm outbreak is initiated by the introduction of five infectious midges. In the fly-model infectious vectors must fly and find hosts. Vectors can either

fly actively or be carried passively on the wind. Both types of movement are described by a random walk Gaussian kernel, where wind is directional according to wind direction. When including vector movement it also becomes imperative to include host distribution onto fields, because the average distance between farms is large compared to the active flight range of midges while fields often have adjoining neighbors. The spread parameters in the fly-model describe the distance over which vectors can locate hosts, the median daily active flight distance, and the time spent carried by the wind if such is present. The fly-model in this way attempts to replicate the biological behaviour of the vectors, when they fly between hosts on pasture and farms.

Initialization of the simulations was done by replicating the assumed route of introduction of the 2007 outbreak in England. Firstly, infectious vectors were carried by wind from the continent and arrived on a farm a little outside Ipswich on the 4 August 2007. In the same manner 6 farms (two in Cambridgeshire, three in Kent, and one in Sussex) received infectious vectors during the period 2-14 October 2007.

Vaccination of host animals was modeled exactly as in Szmaragd et al. 2010 [67], where vaccination reduced the probability of transmitting virus from host to vector and vice versa. The level of protection was a linear function that goes from zero to 100% as a function of time from vaccination day to time of full protection. The time to full protection in sheep was set to 14 days, and 60 days in cattle. The time of vaccination for each county was replicated according to the Defra roll-out plan, which was implemented in 2008. The level of uptake was defined as percentage of holdings having vaccinated all host animals in the holding. We investigated five different vaccination scenarios with 95%, 80%, 50%, 0% and a variable amount of vaccination uptake. The variable vaccination scheme corresponds to a 50/75/95 level of uptake in the free area / surveillance zone / protection zone as defined by Defra in 2007.

Simulations ran from the 4 August 2007 to 3 November 2008 (cf. Szmaragd 2010 [67]). For each scenario 300 outbreaks (defined as any spread from the seven farms infected initially) were generated. The simulation window was chosen so that temperatures and initialization of the outbreak would match the actual incursion of BTV in UK during 2007-2008.

Hourly **temperature** records for 2007 were obtained from the BADC/MIDAS database [92] for the 19 meteorological stations used as inputs to the model. All simulations used the temperature data from 2007 for both years of the simulation, and each herd used temperature records from the nearest meteorological station.

Estimating spread parameters

Parameters for the kernel-model were estimated using data for the 2006 outbreak of bluetongue virus serotype 8 in northern Europe [46]. Because the fly-model relies on flight parameters of vectors, the spatial distribution of host animals on pasture will strongly influence the parameter estimates. However, the distribution of hosts on pasture was not readily available for those countries affected during the 2006 outbreak in northern Europe. Accordingly, the fly-model's spatial transmission parameters were estimated by fitting to data from the 2007 outbreak in GB for the period 4 August to 31 December (see Supplementary Material SM1). Consequently, both models have been fitted to (fly-model) or validated against (kernel-model) data from the same outbreak: that in GB in 2007 (see Supplementary Material SM1, Figure B.5).

Statistical Analysis

The epidemic sizes (cumulative number of affected holdings) were compared using generalized linear models with negative binomial errors and a log link function. Variance of the in epidemic sizes were compared using Levene's test for homogeneity of variances.

References

References are sorted numerically on page 145 and alphabetically on 157.

Acknowledgments

This research was carried out by KG during a three-month visit to The Pirbright Institute. TS and SG were funded by the Department for Environment, Food and Rural Affairs (Defra) (grant code: SE4204). and used facilities funded by the Biotechnology and Biological Sciences Research Council (BBSRC).

Author contributions

Conceived, designed, and produced data from the fly-model: KG. Inputs on optimizing and fitting of the fly-model: LC. Conceived kernel-model: SG. Implementation of and data from kernel-model: TS. Wrote the paper: KG in collaboration with SG. Critically reviewed the paper: TS LC CE.

Additional information

Competing financial interests

SG has received funding from Intervet International b.v. as a consultancy to assess the impact of Bovilis BTV-8 on transmission in the field. The remaining authors declare no competing financial interests.

Supplementary Material SM1 - Estimation and fitting of parameters for the fly-model.

The fly-model introduces five parameters not previously determined (Graesboell et al. 2012 [90]). These are: (i) the area within which midges locate hosts (grid size); (ii) the probability of flying when there are no hosts in an area; (iii) the probability of flying when there are hosts in an area; (iv) the active local flight distance; and (v) the passive flight time in the wind. The collection of data to estimate these parameters has been initiated, but is not yet concluded. Therefore we need to estimate the parameters from existing data and previous outbreaks. Sensitivity analysis indicated that the three first parameters influence primarily the total number of affected hosts, while the last two parameters influence primarily the spatial spread of disease [90]. The area within which midges can locate hosts was assumed to be 500 by 500 meters, which is a very rough estimate. It is, however, similar to the assumption made by Sedda et al. 2012 [59] that the area is a circle with a 400 meter diameter, which was based on data from mosquitoes. No information is available on the probabilities of flying from an area with or without hosts. Accordingly, values for these parameters were chosen solely based on what maximizes the number of infected hosts given that we observe that outbreaks do occur, and, therefore, the parameters must reflect this. The probability of flying when there are no hosts in an area was set to 95% and the probability of flying when there are hosts in an area was set to 5%.

Estimating the active and passive flight parameters for midges was done by fitting a transmission kernel of the form $p(r)=h/(1+r^{\alpha})$ (cf. de Koeijer 2011 [47]) to data on affected and unaffected farms from the outbreak of BTV in UK during 2007 [93]. The parameters h and α were estimated using the maximum likelihood methods described in Boender et al. (2007) [94]. The parameters h and α were also estimated using the same procedure for 5100 simulated epidemics (running from 4 August to 31 December 2007) consisting of 30 replicates of the model using 170 pairs of flight parameters (i.e. active local flight distance and passive flight time in the wind) (see figure B.4a). For each parameter pair the estimates for h and α obtained from the 30 replicates and those from the 2007 outbreak were ranked 1 to 31 according to ascending value of h or α (figure B.4a). The optimal rank for the value estimated from the 2007 outbreak is 16, as there will then be 15 simulated outbreaks with lower values (ranks 1-15) and 15 with higher values (ranks 17-31) for the individual parameters h or α. The rank scores, R_h and R_α, for the two parameters h and α were then combined as $(R_h\text{-}16)^2(R_\alpha\text{-}16)^2$ to determine the best set of flight parameters (figure B.4b). Using rank scores gives equal weight to h and α when combining them, and makes no assumptions about the frequency distribution of the simulated results. Es-

timates for the flight parameters are correlated and the optimal parameter set is not uniquely defined (figure B.4b). To the best of our knowledge, however, and following discussions with entomological colleagues midges are not likely to fly actively more than a few hundred meters per day. Therefore, we chose a local flight distance of 500 meters and a flight time in the wind of 1.5 hours, a parameter combination located in the valley of minimum combined ranks (figure B.4b). In the model the local flight distance is defined as the distance which 50% of midges fly assuming a Gaussian kernel [90]. The flight time in the wind is multiplied by the mean wind speed (sampled uniformly from zero to five m/s) to give the distance which 90% of midges travel assuming a Gaussian kernel [90]. Model predictions for the fly-model using the selected parameters are compared with the predictions of the kernel-model and the observed outbreak in figure B.5.

References

References are found sorted numerically from page 145 and alphabetically from 157.

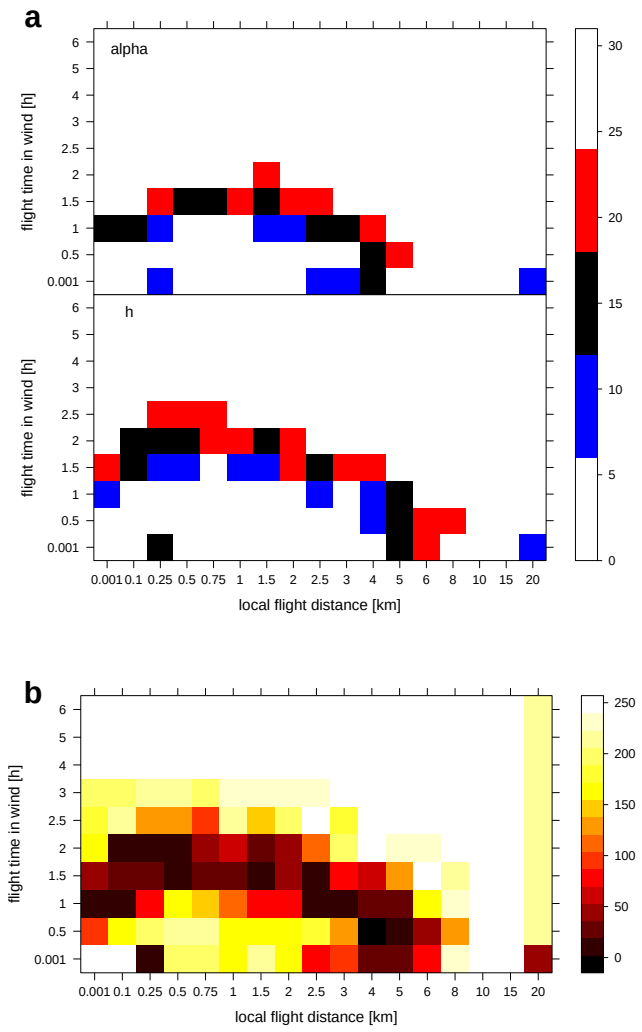

Figure B.4: Rank score plots for the midge flight parameters. The x-axis gives the local flight distance of a midge; y-axis gives the flight time in the wind. (a) The value of h and α estimated from outbreak data is given a rank score compared to the values estimated from the 30 simulations per parameter set, with 16 being the optimal rank. (b) The ranks of subplots of (a) are subtracted 16, squared and multiplied together to determine best parameter value. However there is no single minimum value rather a 'valley' of optimal values indicating that local and wind flight parameters are inversely correlated.

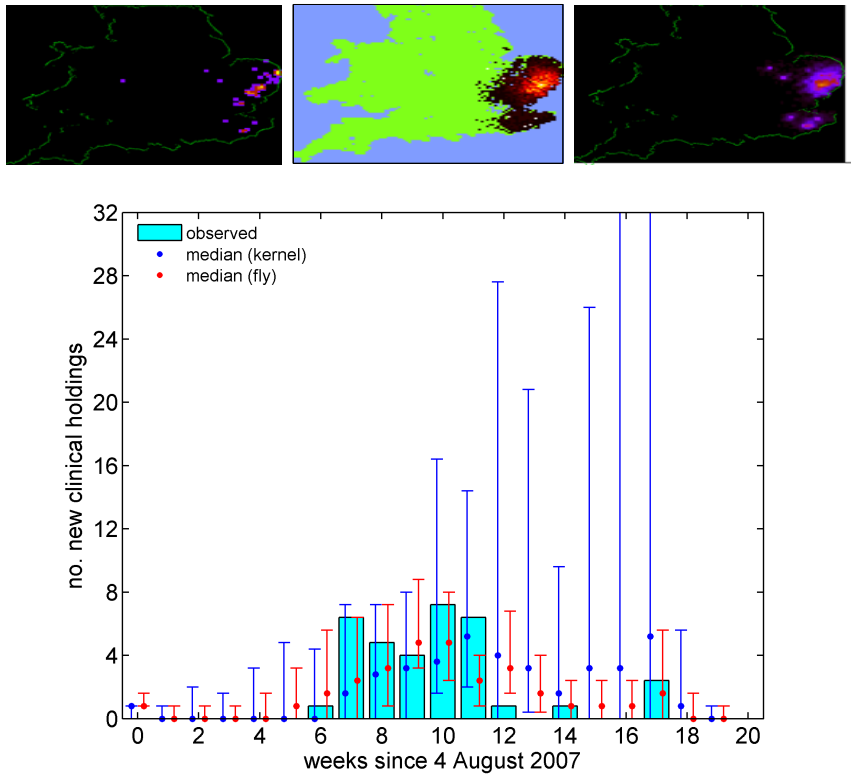

Figure B.5: Outbreak before vaccination. Top left panel: affected holdings during the 2007 outbreak of BTV in England, the color code indicates increasing number of infected farms within a five by five km pixel. Top middle panel: the 2007 outbreak as predicted by the kernel-model, the color code indicates the proportion of simulations that resulted in affected farms within each pixel [46]. Top right panel: outbreak as predicted by fly-model, the color code indicates the proportion of simulations that resulted in affected farms within the pixels. Bottom panel: time course plot of number of new holdings displaying clinical signs for BT per week since 4 August 2007, where boxes indicate the observed cases and circles represent the predicted median number of cases with error bars for the 10th and 90th percentiles for each model.

Appendix C

Scheduling sampling to maximize information about time dependence in experiments with limited resources

---

Kaare Græsbøll, Lasse E. Christiansen.

20 September 2012: Submitted to *Biological Rhythm Research*.

9 October 2012: Accepted for publication.

15 November 2012: Published.

Abstract

Looking for periodicity in sampled data requires that periods (lags) of different length are represented in the sampling plan. We here present a method to assist in planning of temporal studies with sparse resources, which optimizes the number of observed time lags for a fixed amount of samples within a fixed time window given a maximum time lag of interest. The method can also optimize the temporal sampling specifically for situations where samples are at risk of being rescheduled due to otherwise unpredictable events such as weather, faulty equipment, etc. The method is based on the framework of simulated annealing in which we have defined an energy function to be minimized. We compare the calculated sampling plan with a random plan and a cyclic design, and demonstrate how our calculated plan provides the most information about temporal autocorrelation.

Keywords: sparse sampling; autocorrelation; simulated annealing

Introduction

In the epidemiology department of our university we had some old data on trap catches of female biting midges. This data was very sparse, only consisting of 17 samples obtained across a period of 47 days. We wanted to test whether we could obtain the egg-laying cycle of female biting midges from this data, as an estimate of this parameter could be used in a newer project. Biting midges that have laid eggs can be distinguished from those that have not, and are referred to as parous (Dyce 1969) [95]. Therefore our task was to search for correlation in time between the parous midges and the total female population, thereby determining the egg-laying cycle (in this example we will exclude all other confounding factors). However we found that the data had no samples separated by five days (lag 5), and so we were unable to investigate whether there was a correlation in this lag. Unfortunately many lags were poorly represented, and we decided to investigate whether better planning could help avoiding this in the future.

In general, optimal sampling is naturally equidistant sampling at every relevant time unit, be it days, hours, or seconds. Thereby we gain the maximum number

of samples for every lag. But optimal sampling is often not achievable due to various factors, most often these are shortage of time or financial resources. If the hypothesis is that the time dependence is long compared to the sample window, then the solution is to spread the samples equidistantly. If, however, the hypothesis states more complex time dependence, where we expect to see correlation on one or more time periods within the sample window, a good sampling plan is less simple to deduce. We could test all sampling plans to test which gives the most observed lags. This procedure has some limitations because the number of possible permutations in a sampling plan approaches very high values as the number of samples increases. For example: To distribute n=5 samples over N=60 days gives c=30856 possible permutations, and n=20 gives c=$4.5 \cdot 10^{14}$, where $c = (N-2)!/((n-2)!(N-n)!)$, a slightly modified version of the combinatorial to account for two fixed endpoints. It therefore quickly becomes impossible to test all permutations within reasonable time. To overcome this problem of large numbers, we use the simulated annealing algorithm, which can effectively search for the optimum distribution of samples.

A problem when sampling in vivo is that samples might need to be moved once the sampling has begun. There are plenty of examples of how this could happen, such as faulty equipment, disease in personnel, bad weather, etc. When this is the case, samples can either be moved or be cancelled. Our approach can optimize for both these conditions. One of the ways to do this is to plan so that a missing sample can be moved to a future available sample time, and this change in plan will not greatly affect the effectiveness of the sampling plan.

There is little existing literature on maximizing observed time lags with sparse sampling. Most literature deals with the treatment of already obtained data, or how to obtain perfect reassembling of a perfect signal (aliasing). We will therefore compare our work to that of Clinger and Van Ness (1976) [96], which devised how repeated cyclic sampling with a period shorter than the total time period gives samples in all time lags.

Methods and theory

Simulated annealing is a well-tested method and widely used on many optimizing problems (Kirkpatrick *et al.* 1983; Aarts & Laarhoven 1989) [97, 98]. The key aspect is to employ an energy function, E, that describes the system. This energy function must be defined in such a way that it has a minimum energy when the system is in its optimum state. Furthermore, it should be reasonably smooth to be able to guide the algorithm towards the minimum value. A more smooth energy function gives a more effective search for the minimum energy.

We define our energy function as:

$$\begin{aligned} E(\boldsymbol{s},\gamma) &= E'(\boldsymbol{s},\gamma) + \alpha_\rho \rho_{1,2}(E'(\boldsymbol{s},\gamma)) \\ E'(\boldsymbol{s},\gamma) &= \alpha_{\text{mean}}\beta_{\text{mean}}(\boldsymbol{s},\gamma) + \alpha_{\text{min}}\beta_{\text{min}}(\boldsymbol{s},\gamma) + \alpha_{\text{var}}\beta_{\text{var}}(\boldsymbol{s},\gamma) + \\ &\quad \alpha_{\text{gap}}\beta_{\text{gap}}(\boldsymbol{s}) + \alpha_{\text{key}}\beta_{\text{key}}(\boldsymbol{s},\gamma) \end{aligned}$$

Where $\boldsymbol{s}$ is the sampling plan and γ is an input parameter and the only assumption needed about the sampling plan. The sampling plan $\boldsymbol{s}$ is defined by the time period, N, the number of samples, n, and optionally the γ. The αs are given to weight the β and ρ functions, as explained in the following paragraphs, towards a sampling plan that best fulfils the purpose of a given project. To clarify the use of functions we introduce an example which has N=12, n=5, and γ=5: The example sampling plan has samples at times $\boldsymbol{s}$={1, 2, 3, 7, 12}, with the unit of time being determined by the experiment. We also introduce o_k which is the number of observed time differences (lags) which differs by k time.

$$o_k(\boldsymbol{s}) = |\{\boldsymbol{s} - \boldsymbol{s}_i = k \ , \ i = 1, 2, ..., n\}|$$

In the example set, $\boldsymbol{s}$, there is a time difference of one between $\boldsymbol{s}_1$=1 and $\boldsymbol{s}_2$=2, and $\boldsymbol{s}_2$=2 and $\boldsymbol{s}_3$=3, therefore $o_1 = 2$. It follows that o_2=1, o_3=0, o_4=1, etc. From this point on we do not explicitly state that o_k is a function of the sampling plan $\boldsymbol{s}$.

The γ is the maximum lag optimized for in the energy function, where a lag is the difference in time between two samples. If prior to the sampling some information of what might be the largest periodicity in the data is available, this should be reflected in the γ input. For example if it is known that the largest periodicity in the data is approximately 10 time units, the γ unit should be set to this value plus an extra value that reflects the uncertainty of this estimate, e.g. γ=20. It is not necessary to define the γ input parameter, but the lower the γ the faster the program will converge.

$$\beta_{\text{mean}}(\boldsymbol{s},\gamma) = \gamma^{-1}\sum_{k=1}^{\gamma} o_k$$

β_{mean} describes the mean number of observed lags within the maximum lag γ in the sampling plan, $\boldsymbol{s}$. In the example set with γ=5 then mean $\beta_{\text{mean}}(\boldsymbol{s}, 5)$ = (2+1+0+1+2)/5. Notice that if γ is equal to the whole period (12 in the example), mean lag becomes constant no matter how the samples are permutated. Since we would like to have as many possible lags below max lag to observe, mean lag should be maximized by the program, hence α_{mean} should be negative since the technique is to minimize E.

$$\beta_{\text{min}}(\boldsymbol{s},\gamma) = \min(\{o_k \ , \ k = 1, 2, ..., \gamma\})$$

$\beta_{\min}$ gives the minimum count of lags within max lags, γ. In our example o_3=0, and therefore $\beta_{\min}$ becomes equal to zero. The intention is to maximize this value, so that most lags are represented with as high a count in o_k as possible. With the same reasoning as with α_{mean}, $\alpha_{\min}$ should be negative. Together with β_{mean}, the $\beta_{\min}$ drives up the number of samples within max lags, this reduces the significance levels in lag-dependent functions such as the autocorrelation function.

$$\beta_{\text{var}}(\boldsymbol{s}, \gamma) = \text{var}(\{o_k \ , \ k = 1, 2, ..., \gamma\})$$

β_{var} is the variance of the count of lags within max lags, γ. Again, using the example, the number of lags within max lags is {2, 1, 0, 1, 2} and the variance is 0.7. This value must be minimized to ensure that significance levels in the lag-dependent functions are harmonized. As this factor is minimized α_{var} is positive.

$$\beta_{\text{gap}}(\boldsymbol{s}) = \frac{n}{N} \max(\{\boldsymbol{s}_{i+1} - \boldsymbol{s}_i \ , \ i = 1, 2, ..., n-1\})$$

β_{gap} is proportional to the largest time between neighbour samples. Reducing this parameter helps avoid clustering of samples. The β_{mean} and $\beta_{\min}$ terms of the energy function tend to keep all data within γ, which will give a high degree of clustering, usually in the beginning of the time period. In many cases this will be inconvenient as the system may not be stationary, e.g. with cyclic behaviours (generations, recurrence, etc.), or there may be long time trends (e.g. seasonal variations). An example: We could be looking for the egg-laying cycle of an insect, and we expect a cycle of approx. five days, and so set the γ to 20, however the observed period is 60 days; without the β_{gap} function the energy function would cluster a lot of the samples, but often we would like to have more samples throughout the whole period, e.g. to observe generations or to get input at wider range of temperatures. A gap between samples is worse when there are many samples and less so if there are many days to observe in, therefore gap is scaled by the number of samples divided by the sample time span. So in our small example $\beta_{\text{gap}} = 5/12 \cdot 5$. The α_{gap} is positive.

$$\begin{aligned} \beta_{\text{key}}(\boldsymbol{s}, \gamma) \quad = \quad & \max(\{|\{-\gamma \leq \boldsymbol{s} - \boldsymbol{s}_i \leq \gamma\}| \ , \ i = 1, 2, ..., n\}) - \\ & n^{-1} \sum_{i=1}^{n} \{|\{-\gamma \leq \boldsymbol{s} - \boldsymbol{s}_i \leq \gamma\}|\} \end{aligned}$$

β_{key} tests whether a sample is critical for β_{mean} and $\beta_{\min}$ by counting how many times a sample is within max lag of other samples and reporting the difference between max value and the mean value of all samples. Counting lags both forward and backwards regardless of sign in the example, giving {1, 2, 6, 11}, {-1, 1, 5, 10}, {-2, -1, 4, 9}, {-6, -5, -4, 5}, and {-11, -10, -9, -5}, so that sample 1, $\boldsymbol{s}_1$, at time $\boldsymbol{s}_1$=1 has two samples within γ=±5, $\boldsymbol{s}_2$ has three, $\boldsymbol{s}_3$ has three, $\boldsymbol{s}_4$ has three, and $\boldsymbol{s}_5$ has one sample within max lag. In this example, $\boldsymbol{s}_{2-4}$ shares the same value of samples within max lags, and the difference with the mean

$\beta_{\text{key}} = 3$ - (2+3+3+3+1)/5 = 0.6. This part of the energy function thereby does something that is similar to the β_{var}, and when there are only a few samples in $\boldsymbol{s}$ it may be superfluous. However, with a large number of samples the β_{var} term becomes less dependent on a single point, and the β_{key} becomes an effective way to make sure that a single sample does not become too important for the sampling plan. Having such a key sample may be risky for a sampling plan. If a key sample is cancelled, a lot of time lags can be lost. Therefore key focuses on removing such key samples, which minimizes the loss of observed lags in cases of cancellations. The α_{var} is positive because β_{var} should be as small as possible.

$\rho_{1,2}$ is a way to reduce the influence of forced movements of samples. This part of the energy function is computationally demanding, and for this reason should be turned off when moving of samples is not relevant. The method comes in two modes.

ρ_1 takes all samples (except the first and last, since that would change the time period) one at a time and move them to the nearest non-taken time in the future and calculates the energy function, E', for this new sampling plan. Finally, the average over these permutated sampling plans is returned as the value of ρ_1. As the 1 value is the average over energy functions, it should be minimized, and therefore the α_ρ must be positive. Using our example, ρ_1 tests the sampling ensemble of {1, 4, 3, 7, 12}, {1, 2, 4, 7, 12}, and {1, 2, 3, 8, 12} to check that these changes in a sampling plan provides good alternatives to the original plan. If it is not possible to move to a future sample time, the sample is omitted from the sampling plan, and E' is calculated without it.

ρ_2 is more computationally demanding as it checks for the optimal permutation of all samples. It takes every sample (again except the first and last) one at a time, and permutates those to every possible future sample. Using our example $\boldsymbol{s}$={1, 2, 3, 7, 12} for the second sample, $\boldsymbol{s}_2$, ρ_2 checks every available future location {4, 5, 6, 8, 9, 10, 11} and performs the equivalent test for sample $\boldsymbol{s}_3$ and $\boldsymbol{s}_4$. For each sample the minimal value of E' of the permutated sampling plans is determined, and ρ_2 returns the average of these minimal values. Thus we check whether a good sampling plan exists given the movement of any sample.

ρ_1 is devised for situations where samples must follow as close as possible to the originally planned sample plan, whereas ρ_2 is for use when samples can be freely rescheduled after cancelations.

The example used with N=12 and n=5 has only c=792 possible unique permutations, and the ideal solution when γ=5 would be {1, 3, 4, 7, 12} where $E = -6/5 - 1 + 0.16 + 0.1 \cdot 4 \cdot 5/12 - 0.12 - 1.77$, where the αs are normalized to ± 1 except α_{gap}=0.1.

Note that the program is based on a guided random search, so it will not with certainty return the optimal sampling plan, but will return something very close to optimal. In particular, it will not necessarily return the same result when run successively. Therefore it is recommended to run the program more than once to observe how the final value of the energy function fluctuates. If differences in final energy of the system are small, then the program has probably converged towards a good sampling plan. If fluctuations are large the number of iterations in the program should be increased, to allow for better convergence. It is inherent that larger sampling plans require more iterations to converge. However, changing values of αs makes it complicated to determine a good prior estimate on the number of iterations; therefore it is left to the user to specify the number of iterations based on the above procedure. The program automatically checks how many permutations are theoretically possible, and if this number is smaller than the number of iterations, the program will do a full check of all permutations rather than the simulated annealing procedure.

The αs are by default equal to ± 1 except α_{gap}=0.1, where the signs are fixed as stated in the paragraphs relating to the corresponding βs. The αs can be adjusted in order to turn more or less weight to individual parts of the energy function. If it is important for a given experiment that there are no large gaps in the experiment, the value of α_{gap} should be increased. If some part of the energy function should be turned off, the corresponding α value should be set to zero.

We will not describe the functionality of simulated annealing in detail as this can be found elsewhere (e.g. Aarts & Laarhoven (1989) [98]). We briefly state that ergodicity is fulfilled given that we allow permutation to all sample times. Detailed balance is also trivially achieved given that no directionality is expressed by the energy function. We use the Metropolis-Hastings algorithm to evaluate the steps between states.

Cyclic sampling

Clinger and Van Ness (1976) [96] provided a mathematical framework to calculate how to perform cyclic sampling with a period p which is smaller than the entire sampling window N (in principle $p \ll N$) with a minimum number of samples per cycle $f(p)$ and still achieve samples in every lag, i.e. the complete covariance sequence and spectrum in the discrete time (for $N \to \infty$). Since we are only interested in lags smaller than γ this theory is still applicable for finite values of N. The original problem addressed by Clinger and Van Ness was also for a finite value N. We do not recreate their method but have taken sampling plans of different p from Table 1 in their publication (Clinger & Van Ness 1976).

The method does place samples very nicely in every time lag, however given the cyclic nature, the number of observed time lags varies depending on which lag, k, we look at. The original article provides sampling plans for most values of p from 1 to 57. The sampling plan was implemented by starting on day 1 and placing sample times according to the method for p until $n-1$ samples were placed and then one point in the end point at day N. (For example see Figure C.1)

Results

The method we have described had different outcomes depending on the specific sampling plan defined by n, N, and γ. So, here we compare it to the data set we presented in the introduction. The experimental data was n=17 samples over N=47 days and we set γ=20. Female biting midges are assumed to continually lay eggs with the same interval, the egg-laying cycle, from emergence until death. The choice of γ reflects that we expected an egg-laying cycle of 5-10 days. By choosing γ=20 we focus on correlation within 20 time units. The experimental sampling plan had a very inconsistent number of time lags, see Table C.1 and Figures C.1 and C.2a. We observe that our method results in considerably more samples both at an average, β_{mean}, but especially also for the minimum value of samples per lag, β_{min}.

We observe that unforeseen movements of samples during the sample period (Table C.2 and Figure C.3a) still give a higher β_{min} compared to the original conducted experiment, and the total mean of observed lags within max lag, β_{mean}, is at least 20% higher in all instances with moved samples compared to the experiment.

To compare with the cyclic sampling plan we tested for all p in 1-28 and compared the average of observed lags within maximum lag, β_{mean}, and the minimum number of observed lags as a function of lag, β_{min}. In general, we observed that the smaller the p the larger the β_{mean}, but the smaller the β_{min}. The trivial example is p=1 where all samples are next to each other (except the last). This gives a very high mean because almost all lags are below γ (recall that for γ=N all sampling plans have the same β_{mean}). However for p=1 the β_{min}=0 because there are no observed lags for k when $(N-(n-1)) > k > (n-2)$.

For p in 1-28, the β_{min} were never larger than two. We therefore selected the smallest value p with β_{min}=2 which is p=6. As seen in Figure C.1 this allocates no samples in the last third of the sampling period. Therefore we also chose to compare with p=14 which better fills the whole period. The resulting observed

number of lags can be seen in Figure C.2. It is evident that the program sampling plan is superior to both cyclic sampling plans both in β_{mean} and β_{min}. When comparing to Table C.2, we see that even when a sample must be moved the resulting sampling plan is always better than both p=6 and p=14 in β_{min}, and for β_{mean} all are better than p=14, and 9 is better or the same as p=6.

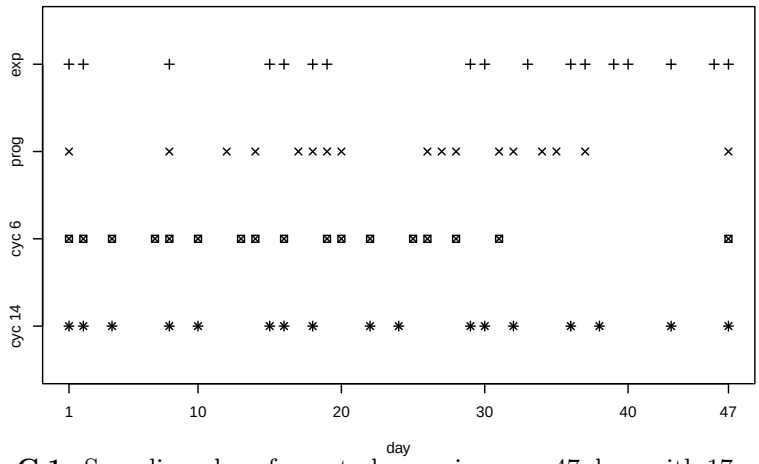

Figure C.1: Sampling plans for a study running over 47 days with 17 samples. The sampling plans are a random plan from a previous experimental study (exp), the plan proposed by our program (prog), and the cyclic plans with period 6 (cyc 6) and 14 (cyc 14) as proposed by Clinger & Van Ness (1976) [96].

lag k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	mean
exp	7	2	8	6	0	5	8	3	2	7	5	1	4	7	3	2	6	5	1	2	4.2
prog	7	6	6	5	5	7	6	6	6	5	5	5	5	5	5	5	5	5	5	5	5.45
cyc 6	5	5	10	4	5	13	4	4	8	3	4	10	3	3	6	3	3	7	3	2	5.25
cyc 14	3	6	3	4	3	8	6	7	4	2	4	4	3	11	3	4	3	3	2	5	4.4

Table C.1: For each of the sample plans presented in Figure C.1 we count how many samples are spaced with k time units between samples, o_k. The rightmost column is the mean of the other columns equivalent to $\beta_{\text{mean}}(\boldsymbol{s}, 20)$. The table values are visualized in Figure C.2.

Comparison using simulated data

To exemplify the difference between our approach and the experimental setup we tested with simulated data from the model of Holmes and Birley (1987) [99]

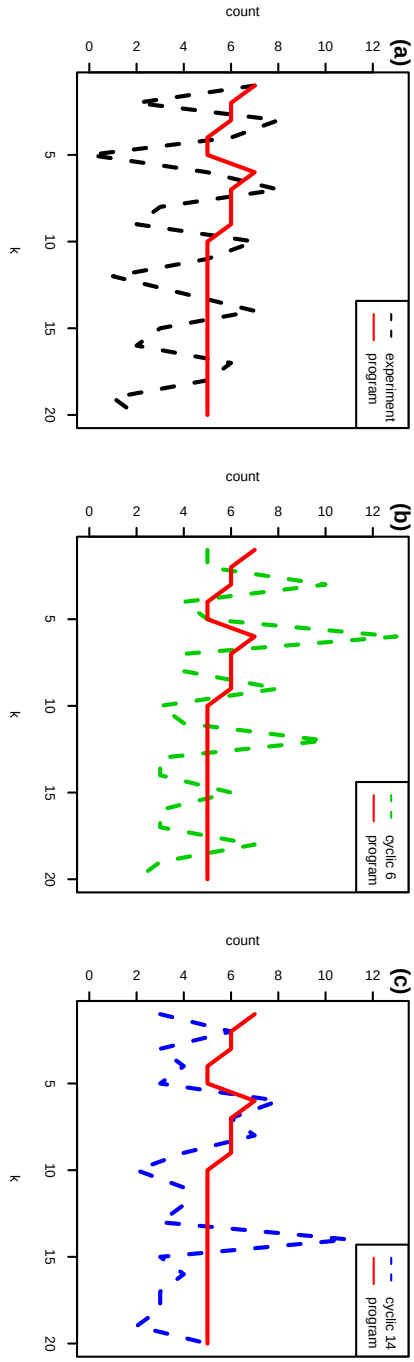

Figure C.2: For each of the sample plans presented in Figure C.1 we count how many samples are spaced with k time units between samples, o_k. The figures compare our method with experiment (a), data spaced cyclically with period p=6 (b), and data spaced cyclically with period p=14 (c). The numbers are presented in Table C.1.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	mean
ii	8	6	7	5	5	7	6	6	6	5	5	5	5	5	5	5	5	5	5	5	5.55
iii	8	6	7	5	5	6	6	5	6	5	5	6	6	5	4	5	6	5	5	5	5.55
iv	7	6	7	5	5	5	7	6	6	5	6	5	4	5	5	6	5	4	6	5	5.50
v	6	5	5	5	4	7	7	6	5	5	4	6	6	4	4	5	5	5	5	4	5.15
vi	5	6	6	4	5	6	6	6	5	4	6	5	5	5	5	4	4	6	5	5	5.15
vii	5	5	7	5	4	7	5	5	6	6	4	5	5	5	4	5	6	4	5	5	5.15
viii	6	5	5	6	5	6	5	5	7	5	5	4	5	4	5	6	5	5	4	5	5.15
ix	6	6	6	6	5	5	6	6	4	5	5	5	5	4	5	5	5	4	6	6	5.25
x	5	6	6	4	6	7	5	5	5	4	6	5	4	6	4	6	5	5	4	4	5.10
xi	7	6	6	6	6	6	6	5	4	5	5	6	6	4	5	5	6	6	4	4	5.40
xii	6	7	4	4	5	6	7	7	6	5	5	5	5	4	5	4	4	5	5	6	5.25
xiii	6	5	5	4	5	6	7	7	6	5	6	4	4	5	4	6	5	4	5	4	5.15
xiv	6	5	5	5	6	6	6	6	7	5	5	6	5	5	4	4	4	5	5	5	5.25
xv	6	5	6	4	5	7	6	5	5	6	5	5	6	5	4	5	5	5	5	5	5.25
xvi	7	6	5	5	4	6	6	6	5	4	5	5	6	6	5	5	5	5	5	4	5.25

Table C.2: From our calculated sampling plan we have moved sample $\boldsymbol{s}_i$ for i in ii-xvi to a future time point (in concordance with ρ_2), in this permutated sampling plan we count how many samples, o_k, are spaced with k time units between samples, and the mean for $k \leq \gamma$. The table values are visualized in Figure C.3a

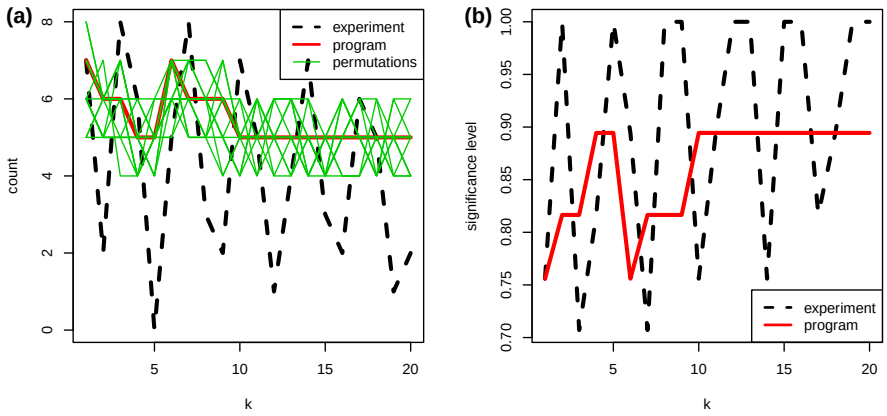

Figure C.3: (a) We take Figure C.2a and overlay the 15 permutations of sampling presented in Table C.2. (b) The significance level of the autocorrelation function, $2/\sqrt{o_k}$, given the number of observed lags ok for the k'th lag. Meaning that observed autocorrelations are only significantly different from zero on an approximate 95% level if they exceed these values, e.g. to detect a true autocorrelation of 0.5, o_k should be larger than 16.

how well we can determine the egg-laying cycle of biting midges. The simulated model was given as:

$$\begin{aligned} P_t &= S \cdot T_{t-u} \\ N_t &= U[0; R] \\ T_t &= N_t + P_t \end{aligned}$$

Where T_t is the total population of female biting insects comprised of N_t, nulliparous (which denotes midges that have not laid eggs), and P_t, parous (which denotes females that have survived their first egg-laying cycle), all as a function of time, t. The parameters are u, which is the length of the egg-laying cycle, S=0.50, the probability of surviving a cycle, and R=1000, the maximum recruitment rate of new nulliparous, so nulliparous is drawn each day from a uniform random distribution between zero and R, $U[0; R]$. We have included an example of simulated data (Figure C.4), and we should point out that only 17 samples are taken from the simulated data to determine the parameter u.

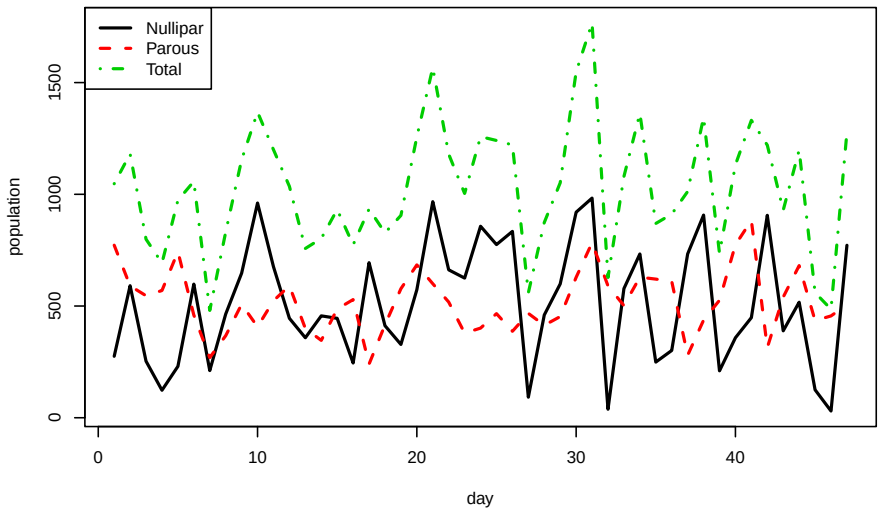

Figure C.4: Example of simulated data of population dynamics in biting midges with egg-laying cycle u=10. From this population data 17 samples are taken according to the different presented sampling plans (Figure C.1) to determine u.

We run this model 10000 times for each integer u between 1 and 20, to test

how often the correct egg-laying cycle can be located as significant in the autocorrelation function, where the significance level is $2/o_k^{1/2}$ given that o_k is the number of paired samples (lags) for the time distance k in the sampling plan (Table C.1 and Figure C.3). The significance level puts an upper limit of 50% positively identified significant egg-laying cycles in the conducted experiment, as only 10 lags had more than the four occurrences required to go below 1 in the significance level. This is also clearly reflected in the results: Using the conducted experimental sampling plan 47.9% ± 0.5% simulations had the correct egg-laying cycle as significant, while this was so for 92.6% ± 0.9%, had the samples been done according to the plan proposed by our program. With the same amount of samples our chance of determining the correct egg-laying cycle was doubled. For the cyclic sampling plans the theoretical upper limit of number of significance in the autocorrelation function would be 45% and 30% with period 6 and 14 respectively.

In this simulation example, our program comes out at least two times better than all other options. This is largely due to the very limited number of samples, n, and the consequent balancing on the significance level in the autocorrelation, which has the critical limit at four samples per lag k. Had n been larger the advantage of our program would not be as large, given that the significance level scales with the square root of o_k. However, when having real data, every count counts, and the model is primarily designed with robust sparse sampling in mind.

Conclusion and discussion

We have devised a method for planning temporal experiments in situations where limited resources prevent testing at all available time slots. The sampling plan suggested by our method is superior to the sampling plans it was compared with. It is also very robust to changes in sampling plan, and when sampling times are forced to change, it is still better or almost as good as the default cyclic sampling plans it was compared with (depending on which β the comparison is done with).

The result of the program depends greatly on the variables α. Therefore we have not provided a table of results, but rather provided an example of use. A general observation about the sampling plans with normalized αs is that they have a tendency to create one cluster of many samples on the scale of max lag (γ) and then scatter the remaining over the full period. From testing it was evident that certain combinations of values for the αs limit the influence of the permutation functions, ρ. Therefore it is our advice that when tuning the values

of αs to achieve a sampling plan best suited to a specific experiment, one should also try turning $\rho_{1,2}$ on and off.

The method we have developed was written into a package in R. However the underlying code was written in Fortran 90, whereas CRAN only supports Fortran 77. We can therefore only publish this package ourselves and it will only run on Unix/Linux systems which feature a Fortran 90 compiler (which the majority of Unix/Linux systems do). Please email the corresponding author for a copy of the program.

The type of planning presented in this paper is useful for all experiments where the temporal aspect is desired to be analysed and correlations in time are expected to be found. The approach is also valid for all spatial experiments when the spatial sampling pattern is unchanged in time. Moreover, this method can be utilized to determine the number of samples, n, needed given a desired level of significance in the autocorrelation function. In the previous example, we could have removed one sample and the calculated sampling plan would still yield better results than the 17 samples in the actual experiment. If it is the goal to obtain as high a value as possible of observed lags, o_k, for all k, it seems tempting to set all αs but $\alpha_{\min}$ to zero. We do not recommend this procedure. Only focusing on one α, no matter which one, will dramatically reduce the smoothness of the energy function E as a function of $\boldsymbol{s}$. This is undesirable since the simulated annealing depends on local gradients towards the minima, as do most optimizing procedures.

Acknowledgements

We thank Peter Lind and Claes Enøe for a thorough read of the article and several good comments on how to improve readability for the biologically minded. We also wish to thank Carsten Kirkeby for allowing us to use his data as the example.

References

References are found sorted numerically from page 145 and alphabetically from 157.

Author contributions

KG conceived the idea and the method, wrote all the code, and wrote the paper. LC gave inputs to the main equation and critically reviewed the paper.

Appendix D

The range of attraction for light traps catching biting midges (Diptera: Ceratopogonidae: Culicoides) vectors

---

Carsten Kirkeby†, Kaare Græsbøll†, Anders Stockmarr, Lasse E. Christiansen, René Bødker.

† This paper has shared first authorship.

12 September 2012: Submitted to *Parasites and Vectors*.

20 February 2013: Accepted for publication.

15 March 2013: Published.

Abstract

Background: *Culicoides* are vectors of e.g. bluetongue virus and Schmallenberg virus in northern Europe. Light trapping is an important tool for detecting the presence and quantifying the abundance of vectors in the field. Until now, few studies have investigated the range of attraction of light traps.

Methods: Here we test a previously described mathematical model (Model I) and two novel models for the attraction of vectors to light traps (Model II and III). In Model I, *Culicoides* fly to the nearest trap from within a fixed range of attraction. In Model II *Culicoides* fly towards areas with greater light intensity, and in Model III *Culicoides* evaluate light sources in the field of view and fly towards the strongest. Model II and III incorporated the directionally dependent light field created around light traps with fluorescent light tubes. All three models were fitted to light trap collections obtained from two novel experimental setups in the field where traps were placed in different configurations.

Results: Results showed that overlapping ranges of attraction of neighboring traps extended the shared range of attraction. Model I did not fit data from any of the experimental setups. Model II could only fit data from one of the setups, while Model III fitted data from both experimental setups.

Conclusions: The model with the best fit, Model III, indicates that *Culicoides* continuously evaluate the light source direction and intensity. The maximum range of attraction of a single 4W CDC light trap was estimated to be approximately 15.25 meters. The attraction towards light traps is different from the attraction to host animals and thus light trap catches may not represent the vector species and numbers attracted to hosts.

Keywords: *Culicoides*, range of attraction, vector abundance, light traps, vector monitoring.

Introduction

Biting midges (Diptera: Ceratopogonidae: *Culicoides*) are vectors of e.g. Bluetongue virus [30] and the newly discovered Schmallenberg virus in northern

europe [34]. Due to their crepuscular activity pattern, the standard trapping method is by (UV) light traps [100]. The vision of *Culicoides* and Ceratopogonids in general has not been studied well [101], although their phototactic behavior is of epidemiological importance. This behavior influences their response to light traps, which are widely used for determining the presence, abundance and phenology of *Culicoides* (e.g. [102–104]).

An optimal sampling strategy to estimate insect abundance must rely on knowledge of the area or range covered by a single trap. Many different terms have been used for this measure, and we prefer to use the 'range of attraction', describing the (maximum) distance at which insects are attracted to the trap. This term allows for a non-symmetrical attraction range around the trap which is relevant for traps equipped with light tubes. Few studies have attempted to estimate the range of attraction for light traps: Odetoyinbo [105] carried out a study where a trap was hung at different distances from an open window which mosquitoes passed through at night. The aim was to estimate the point where the trap caught more than a simultaneously operated independent trap. Here, the 'effective range' for a CDC mini light trap was estimated to be approximately 5m. Baker & Sadovy [106] put up 125W mercury vapor lamps at different distances from the release point of marked moths. By varying the distance from the release point to the lamps, they found that the response distance was within 3m. More recently, Truxa & Fiedler (2012) [107] carried out a study where marked moths were released at different distances to a UV-light trap. The catches showed that the 'radius of attraction' was up to 40m in 5-min intervals. For *Culicoides*, Rigot & Gilbert (2012) [108] compared 8W Onderstepoort light traps and how they competed with each other at different distances. Background fluctuations in the *Culicoides* abundance were monitored by an independent light trap. The analysis assumed a fixed radius of each trap, regardless of the distance to other traps. The 'effective trap radius' was found to be approximately 30 m when the traps were running for 30-min intervals. However, Venter *et al.* (2012) [109], only found a 'range of attraction' for *Culicoides* for the same trap type of between 2 and 4 meters. In that experiment, two traps were hung at varying distances to each other and the background fluctuations were monitored using an independent trap. The 'range of attraction' was then the distance at which the two traps began to catch less than the independent trap. Both of these last studies hypothesized that *Culicoides* are attracted to the nearest trap, and that the light from the trap is isotropic (uniform in all directions). In the present study we incorporated the directionally dependent (anisotropic) light created by the light tubes in the light traps into two novel models for the attraction of *Culicoides* to light traps.

We tested three different models for attraction of *Culicoides* to light traps: Model I where the range of attraction for each trap is isotropic and independent of the distance to adjacent traps and where *Culicoides* fly to the nearest trap;

Model II where *Culicoides* fly towards the direction with highest light-intensity, simulating anisotropic light and where overlapping ranges of attraction create an extended range of attraction; and Model III where overlapping ranges of attraction also create an extended range of attraction but where *Culicoides* continuously evaluate each light source in its field of view and fly towards the highest light-intensity. We used two different experimental setups to test the models. By fitting the models to the relative trap catches in the experimental setups we exclude factors influencing the level of abundance. The range of attraction for *Culicoides* is then estimated from the model that best fit the trap data in both experimental setups.

Methods and Materials

Experiments were conducted in the summer of 2011, on a farm with approximately 70 cows in Klippinge, Denmark (geographical coordinates: N55.3619, E12.3234). The study field, measuring approximately 120 by 120 meters, was grazed by cattle during the day. Before dusk, the cattle were excluded from the field, but had access to enclosures on the western side of the study site. The surrounding land cover was grazed fields and grain fields, so there were no obstructions of *Culicoides* vision or flight next to the setup. Approximately 100 meters from the experimental setup there was a cow stable and a dunghill with potential breeding sites for the Obsoletus group. In a radius of 500 m, there were at least three ponds with potential breeding sites of the Pulicaris group. However, no breeding sites were monitored during the study period. The experiment was set up close to the cattle to ensure a high density of *Culicoides*. There were no other sources of light pollution present on the field during the experiments. *Culicoides* were caught using CDC 1212 mini light traps (www.johnwhock.com). These traps are equipped with a horizontally mounted 11 cm fluorescent tube emitting anisotropic UV-light. This means that the highest light-intensity is seen perpendicular to the tube and no light is seen from the ends of the tube. The light tubes were placed at a height of 180 cm and all light tubes were aligned along the transect line. Before each catch night, freshly charged batteries were installed on the traps. The starting time of sampling was decided each catch night to be when it was dark enough to perceive the light from the traps with the naked eye at a few meters distance. Traps were allowed to catch in intervals of one hour before they were emptied. Catch nights were chosen subjectively for optimal flight conditions for *Culicoides*: low wind speed; no precipitation; high air humidity; no fog in the air; and not too low temperature. Weather variables were monitored during the experiment using a Davis Vantage Pro 2 weather station.

Experimental setup

Experimental setup A In the first experimental setup, 10 traps were set up in each of two transects with higher trap density towards the middle of the transects, aligned east-west (Fig. D.1). Within each transect, the traps were positioned at 0, 3, 9, 21 and 45 meters from the middle of the transect. In this way, the distance between traps was doubled for each transect position. In the middle position, two traps were placed with the light tubes separated by 12 cm. Two parallel transects were set up in each catch interval and separated by 100 meters. Setup A was run on the 27th (between 22.15 and 00.15 hours), 30th (21.15-00.15 hours) and 31st (22.15-00.15 hours) of July 2011.

Experimental setup B In the second experimental setup, 6 traps were placed in two transects with higher density towards one end, also separated by 100 meters. This setup was either aligned north-east to south-west (E-W) or north-east to south-west (N-S) (Fig. D.1) in each catch interval. Within each transect, the traps were placed at 0, 3, 9, 21, 45 and 93 meters from the starting point, also doubling the distance between traps for each position. The transect directions were reversed so the end with more traps pointed in opposite directions. This setup was run on the 17th (N-S, 21.45-00.45 hours), the 18th (N-S, 21.45-00.45 hours) and on the 24th (E-W, 21.15-00.15 hours) of July 2011.

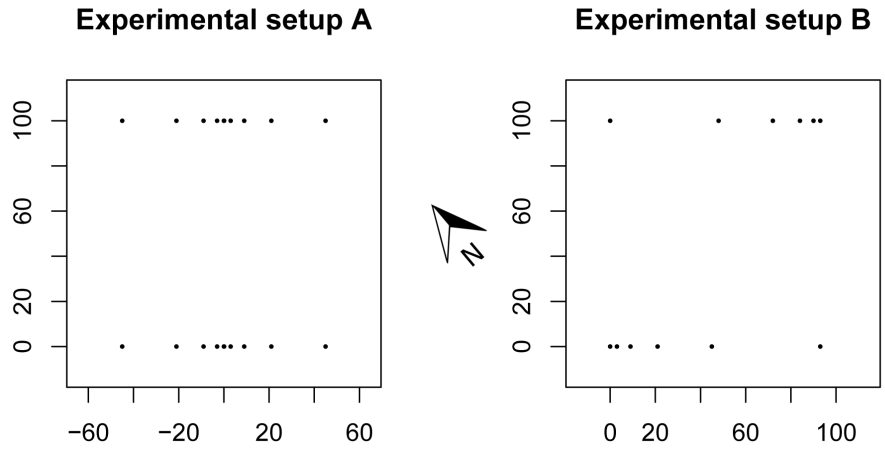

Figure D.1: Experimental setup. Experimental setup A and B. The plots show the study area viewed from above. Distance units are meters. Traps are represented by black dots. All traps were hung with the light tubes along the transect line. In setup A, the middle dot represents two traps separated by 12 cm. In setup B, the configuration was rotated 90 degrees in some time intervals.

All light traps were equipped with new light tubes and were aligned so that all the tubes were parallel to the transect direction. Insects caught in the traps were sorted in a dissection microscope. *Culicoides* were identified by wing morphology according to Campbell & Pelham-Clinton [110] and female specimens of the *Culicoides obsoletus* group and the *C. pulicaris* group were identified and used for analysis. The two species groups contributed separately to the dataset so that the total number of study units were initially 54 transects, consisting of different species groups, hourly catch intervals and transect positions.

Models

We investigated three mathematical models to explain the observed fraction of catch per trap per transect. Model I, 'Nearest trap', assumed a constant trap radius where vectors always fly to the nearest trap, as suggested by Rigot & Gilbert (2012) [108]. Model II, 'Indirect light', calculates the combined light field surrounding the traps, then assumes that *Culicoides* always fly toward areas of higher light intensity, and in this way range of attraction becomes defined by a cutoff value in total light intensity. Model III, 'Perceived light', assumes that *Culicoides* fly in the direction of what they perceive as the brightest light source. The model approximates the perception of a light trap for *Culicoides* with a Gaussian function, which means that the lights from closely placed traps overlap. Range of attraction is again defined by a cutoff in light intensity.

Model I - Nearest trap: Effectively this model states that *Culicoides* always fly towards the nearest trap. Consequently traps are assumed to catch a number of *Culicoides* proportional to the area within the range of attraction, r (IA/B in Fig. D.2). If the distance, d, between two neighboring traps is smaller than two times the range of attraction, each trap's area of catch is reduced by half of the overlapping area. The model only allows for the catch area to be reduced by the nearest neighbor(s). The equations we present are only valid for traps on a line. The predicted fraction of catch, C_i^I, for the i'th trap is then:

$$\begin{aligned} A_i^I &= \pi r^2 - \sum_{j \in NN} \left(2r^2 \arctan\left(\frac{2\Omega}{d_{ij}}\right) - d_{ij}\Omega \right) \\ \Omega &= \sqrt{r^2 - \left(\frac{d_{ij}}{2}\right)^2} \\ C_i^I &= \frac{A_i^I}{\sum_{i=1}^{n} A_i^I} \end{aligned} \tag{D.1}$$

Where A_i^I is the area within range of attraction for the i'th trap minus half the area of the j'th trap, which is the one or two next neighbors, NN, d_{ij} is

the distance between the i'th and j'th trap, and n is the total number of traps in the transect. A_i^I is represented by one color per trap (i) within the white lines (range of attraction) in Fig. D.2. This type of model was investigated for *Culicoides* by Rigot & Gilbert (2012) [108].

Model II - Indirect light: In this model we assume that *Culicoides* always fly toward areas of higher light intensity, quite similar to moving towards higher concentrations of scent molecules, when searching by smell. We therefore calculate the total light intensity, $I(x, y)$, for the area around the traps:

$$I(x,y) = \sum_{i=1}^{n} \frac{\Psi(\phi_i(x,y))}{(x-x_i)^2 + (y-y_i)^2} \tag{D.2}$$

$$\Psi(\phi_i(x,y)) = |\cos(\phi_i(x,y))| \tag{D.3}$$

Where n is the number of traps and x_i, y_i are the spatial coordinates of the i'th trap. *Culicoides* in any given position are then assumed to fly along the highest gradient of $I(x, y)$. We model the anisotropic light field around traps using Ψ, which is a function of the angle around each trap, $\phi_i(x, y)$, in concordance with Lambert's cosine law. To model the attraction area we start one *Culicoides* for every square meter within at least 150 meters from the central trap, and simulate the attraction of light by individually moving them in small steps along the largest gradient in the surrounding light field until they eventually arrive at a trap location. From this we can determine a function $T^{II}(x, y)$ which tells which trap a single *Culicoides* will fly to as a function of initial position (IIA/B in Fig. D.2). We can then define a cutoff, I_C, in light intensity that defines how far away *Culicoides* are attracted towards the light. And from this determine the fraction of catch of *Culicoides* for the i'th trap, C_i^{II}:

$$C_i^{II} = \frac{\sum_{x,y} \mathbf{I}\,(\,I(x,y) > I_C \wedge T^{II}(x,y) = i)}{\sum_{x,y} \mathbf{I}\,(\,I(x,y) > I_C)} \tag{D.4}$$

Where $\mathbf{I}$ is the indicator function. This equation and figure D.2 is then interpreted as that each trap in the transect catch a fraction of *Culicoides* proportional to the area within the light cutoff.

Model III - Perceived light: This model aims to recreate the view that *Culicoides* have of the traps from every point on the area around the traps. *Culicoides* will fly in the direction of the perceived brightest light. For every point on the field a 360° view is generated with a resolution of 1°, $V(x, y, \phi)$. This is calculated by combining the light intensity from the i'th trap at every point, $I_i(x, y)$. Combined with anisotropic light, Ψ, and the inaccuracy *Culi-*

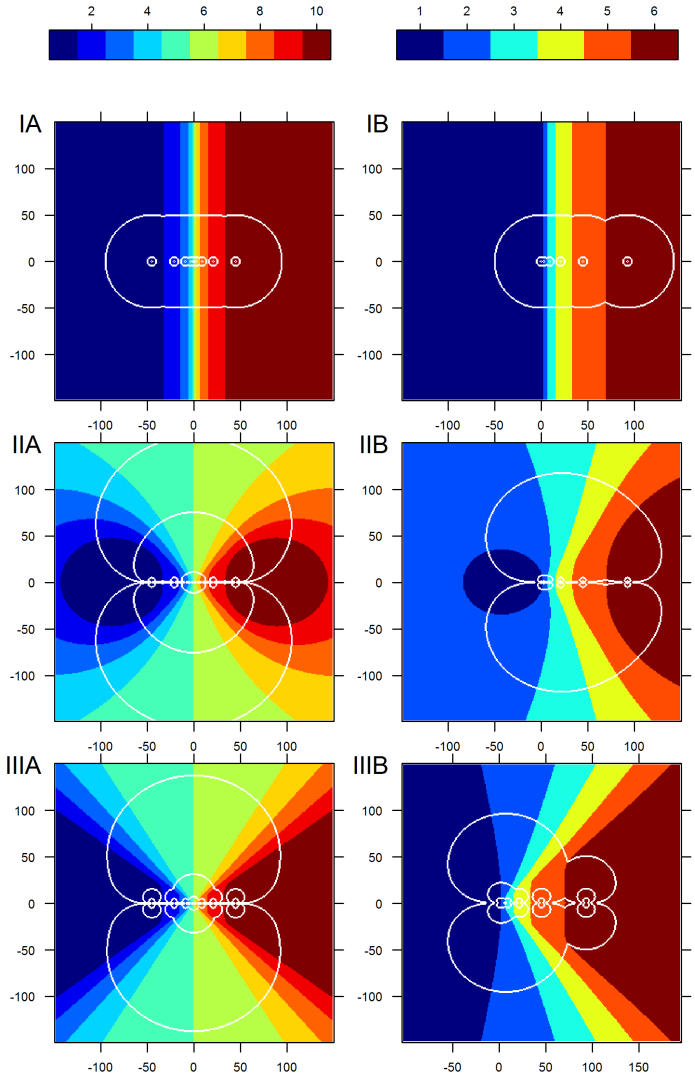

Figure D.2: Range of attraction vizualisation. The area of attraction for the three models. It is assumed that each trap catches a number of *Culicoides* proportional to the area within the range of attraction (white lines). The plots represents modeled fields of 300 by 300 meters with one transect of traps from Fig. D.1. The colors (color-key on top) indicate which trap a *Culicoides* in this area will end up in, with trap numbers corresponding to numbers in Fig. D.5. The white lines indicate different cutoff points in range of attraction (light intensity). These plots thereby represent the functions $T^k(x, y)$, where the white lines in each plot indicate the three cutoffs, I_C, also indicated in Fig. D.5, which corresponds to a range of attraction of 5 and 50 meters and the best fit. Model III is presented with $\sigma = 10$. The left column is setup A, the right column is setup B. In model I (top) *Culicoides* always fly to the nearest trap, in model II (middle) *Culicoides* fly towards the area of highest light intensity, and in model III (bottom) *Culicoides* fly towards what they perceive as the brightest light, as illustrated in Fig. D.3.

coides perceive the position of the light traps, σ.

$$I_i(x,y) = \frac{\Psi(\phi_i(x,y))}{(x-x_i)^2+(y-y_i)^2} \tag{D.5}$$

$$V(x,y,\phi) = \sum_{i=1}^{n} \frac{I_i(x,y)}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\phi-\phi_i(x,y))^2}{2\sigma^2}\right) \tag{D.6}$$

We again start one *Culicoides* for every square meter and simulate their flight towards what they perceive as the strongest light until they arrive at a trap location. From this the function $T^{III}(x,y)$ is determined which describes which trap a single *Culicoides* will fly to as a function of initial position (Fig. D.2). We define a cutoff, I_C, in light intensity that defines how far away *Culicoides* are attracted towards the light. And from this we determine the fractional catch of *Culicoides* per transect:

$$C_i^{III} = \frac{\sum_{x,y} \mathbf{I}\left(\max(V(x,y,\phi)) > I_C \wedge T^{III}(x,y) = i\right)}{\sum_{x,y} \mathbf{I}\left(\max(V(x,y,\phi)) > I_C\right)} \tag{D.7}$$

The *Culicoides* fly in the direction they perceive to have the brightest light. When *Culicoides* are far away the trap lights seem to blend together because the angular distance between them is smaller and the σ smears the view. As illustrated in fig. D.3. **Range of attraction**, r^k, for model II and III must be calculated from the light intensity cutoff, I_C^k, which is a result from the best fit of models to data. In the experimental setups the combination of light from the traps provides a complex pattern for the combined range of attraction as seen in figure D.2. But for a single trap I_C^k determines the range of attraction perpendicular to the tube as:

$$r^k = \sqrt{\frac{\kappa^k}{I_C^k}} \tag{D.8}$$

$$\kappa^{II} = 1 \quad , \quad \kappa^{III} = \frac{1}{\sigma\sqrt{2\pi}}$$

With κ^k being the light intensity one meter from a trap in the k'th model. Model II and III are normalized differently because the total light intensity from one lamp across the view function, V, is set to sum to one in model III.

Fitting to data: To determine best fit we used the value of a χ^2-test statistic (CS^k) to evaluate the modeled fraction of catch C_i^k, from the k'th model, with the observed data, $C_{i,j}$.

$$\mathrm{CS}^k = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{(C_{i,j}-E_{i,j}^k)^2}{E_{i,j}^k} \tag{D.9}$$

$$E_{i,j}^k = C_i^k(r^k,\sigma) \times \hat{\mathrm{I}}_j$$

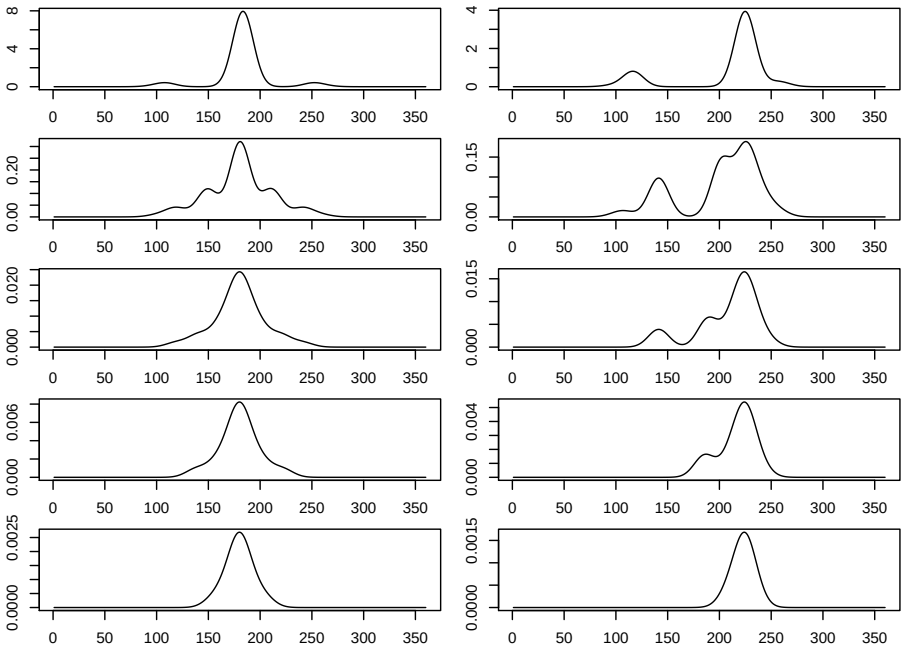

Figure D.3: 360 degrees view of a biting midge. The light intensity of setup A as perceived by one *Culicoides* as a function of degrees on the angle of the transect. X-axes show the view angle in degrees where zero degrees is downwards orthogonal on the transect line and the angle increases counterclockwise. Plots show views as if a single *Culicoides* approaches the central traps in a straight line perpendicular to (left), or in a 45 degree angle to the transect (right). In the coordinate system of Fig. D.2 and equation (D.6), this corresponds to the coordinates $(0, -i)$ (left) and $(-i, -i)$ (right) where $i = 1, 5, 25, 50, 100$ meters, with $\sigma = 10$. *Culicoides* are assumed to fly towards the brightest perceived light, and as the perception is assumed to be Gaussian, *Culicoides* must be close to the transect (distance depends on σ) to differentiate neighboring traps.

Where n is the number of traps, m is the catch number with $\hat{\mathrm{I}}_j$ being an identity vector of length j, so $E^k_{i,j}$ is the expected fraction of catch from model k repeated j times equal to the number of separate catches. And C^k_i is the fractional catch from the models, which is dependent on range of attraction and also σ in model III. The best fit is the set of parameters (r^k, σ) that minimizes the value of CS^k. This method puts equal weight on each transect of catch. Given that the $C_{i,j}$ is the relative catch per trap per transect, the abundance of *Culicoides* does not have any impact on the analysis. This removes the need to include factors which affect abundance in the model.

All of the data was not included in the trap data. A transect of trap data was omitted if there were more than three zero catches, which was usually observed on days with a very low total catch. Many zero catches would bias the catch distribution towards equal catches between traps, which would not represent the true catch distribution, but merely reflect the variation in daily catch. In total 14 of 28 trap data sets for setup A were excluded, while only 5 of 26 where omitted for setup B. Thus a total of 35 transects were included in the analysis. The analysed data from setup A was consequently from the 27th (between 23.15 and 00.15 hours), the 30th (23.15-00.15 hours) and the 31st (22.15-00.15 hours) of July 2011. For setup B analysed data was from the 17th (N-S, 23.45-00.45 hours), the 18th (N-S, 21.45-00.45 hours) and the 24th (E-W, 22.15-00.15 hours) of July 2011.

In the data there were three missing data points (NAs) for setup A and two NAs for setup B. These were handled by using an Expectation Maximization (EM) procedure [111]. The EM converged in all cases after a maximum of one step.

In setup A data is presented symmetrized by averaging over traps in pairs around the center of one transect of traps. We symmetrized data to remove directional bias in the experimental setup. However, the symmetrizing did not affect fitting with the CS function, and symmetrized and un-symmetrized data gave the same results. We chose to present the data symmetrized to allow for a better visual comparison with the models, which will always give a symmetrical result for setup A.

Confidence intervals on r and σ in model III were determined using Fischer information theory as presented in Madsen (2008) [112]. The method was implemented by approximating the CS-test curve to a second order polynomial using a power transformation to symmetrize around the minimum value of the CS function (9). Notice that this method produces non-symmetric confidence intervals.

To ensure that the model was not driven by the catch on one night, one experimental type of setup, or one *Culicoides* species we used the jackknife method on the data. Which is to reanalyze the data excluding the data from one catch night at the time, each species group, or each experimental setup at the time.

Results

10,150 *Culicoides* were caught and included in the analysis, of which 1,817 specimens were from the *C. obsoletus* group and 8,333 specimens were from the *C. pulicaris* group. The hourly catches from each transect ranged between 3.6-27.8 (mean: 13.0) specimens per trap for the *C. obsoletus* group and 2.8-177.8 (mean:52.5) specimens per trap for the *C. pulicaris* group. Each transect in setup A comprised 90-278 (mean: 179.1) specimens from the *C. obsoletus* group and 312-1778 (mean: 970.3) specimens from the *C. pulicaris* group. In setup B, each transect comprised 22-144 (mean: 61.8) specimens from the *C. obsoletus* group and 17-112 (mean: 63.4) specimens from the *C. pulicaris* group.

The catch nights were chosen subjectively for optimal flight conditions for *Culicoides*. During the catches, the temperature was between 12.9 and 18.6 degrees Celcius. The dew point temperature was below the ambient temperature during the whole study, and the air humidity was between 72% and 94%. The wind speed was between 0 and 1.8 m/s, and no precipitation was measured. Thus there was no rain, no fog in the air, high humidity, low wind speed and not too low temperatures.

The best fit of the models was determined by minimizing (D.9) as a function of the range of attraction r for model I, as a function of light intensity cutoff I_C in model II, and as a function of light intensity cutoff I_C and σ in model III (Fig. D.4). For model II and III, I_C is recalculated to r by using (D.8).

The catch distribution of different ranges of attraction and the best fits of the models are presented in Fig. D.5. The collected trap data are numbered as presented in Fig. D.1. from left to right at $y = 0$. We generally observed that traps catch a lower fraction of *Culicoides* when placed closer together. However, there are two very clear exceptions when traps are placed close together. The two central traps in setup A (number 5 and 6) catch almost the same as the outermost traps, and trap 1 in setup B caught a higher fraction than the other traps. These observations are strong indications that closely placed traps do not only compete for *Culicoides*, but also amplify attraction. The characteristics of the models compared to data are as follow (as seen in Fig. D.5).

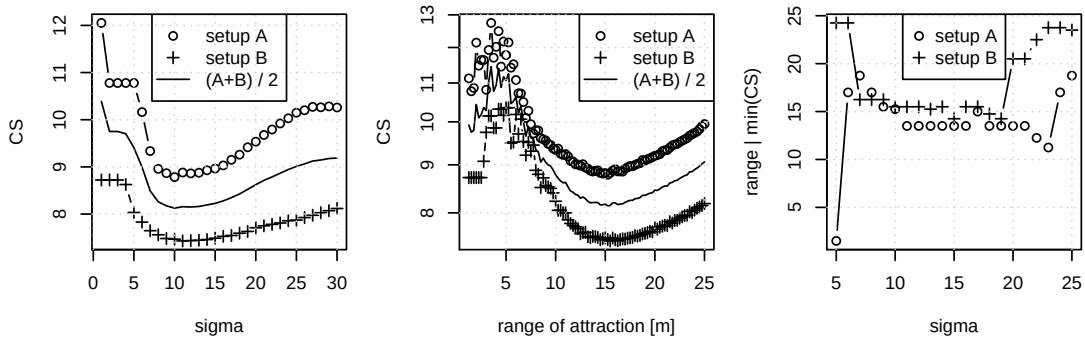

Figure D.4: Model fits of sigma and range of attraction. CS values from eq. (D.9) as a function of σ (left), and range of attraction, r, with $\sigma = 10$ (middle). The minimum CS value indicates which value best fits with the observed data, when using model III. Right is the range of attraction, r, that minimizes CS for different values of σ, which shows a stable range of attraction for a range of σ. All plots are for model III. The jump in values around $\sigma = 20$ in the figure furthest to the right is due to range of attraction exceeding the simulation box size. The noise for small trap radius and small σs is due to rounding errors that occur on small scales.

Model I - Nearest trap: When range of attraction is lower than half the distance between the closest traps, the traps do not compete and will catch the same fraction of *Culicoides*. When going towards higher trap radius the center traps in setup A will catch a lower and lower fraction due to competition with neighboring traps. The outermost traps will always catch the highest fraction due to the lowest competition. Model I is therefore unable to reproduce the remarkable peak in the middle traps which is observed in data from experimental setup A. Moreover, it overestimates the catch in the outermost traps in setup B.

Model II - Indirect light: In this model *Culicoides* are attracted towards higher total light intensity, which is a very simple way of considering attraction to light. Flying towards the strongest concentration of light attracts more *Culicoides* to the central area of the trap setups. This explains that for setup B the model predicts that traps 2 and 3 for medium and large ranges of attraction catch more than the outermost traps, which is contrary to the observed data. We therefore observe that the fitted r^{II} is very different whether fitted to setup A or B.

Model III - Perceived light: *Culicoides* in this model will fly directly towards the perceived brightest light source. The two central traps in setup A are located within such a small distance that *Culicoides* cannot distinguish them until within a very short distance, therefore they will appear as one trap with twice

the brightness (Fig. D.3). This gives the added attraction to the middle traps which produce the central spike in the trap catch distribution, which is also observed in data from setup A. Model III also fits data setup B where competition among traps 2, 3, and 4 reduce their fraction of catch compared to the outermost traps. Moreover, trap 1 is predicted to catch a higher fraction due to to the nearness of trap 2.

In model III the CS function for both setup A and setup B displayed a global minimum at $\sigma = 10$ (left in Fig. D.4). With a combined 95% confidence interval 7.2-13.8. For $\sigma = 10$ the CS is a continuous function of r with a unique global minimum at $r^{III} = 15.25$ meters for both setup A and B (middle in Fig. D.4). With a 95% confidence interval 12.7-18.3. Furthermore, we observed that for a broad range of σs (also covering the 95% confidence interval) the optimal single range of attraction is approximately the same for setup A and B (right in Fig. D.4). Even though r^{III} is reported to 15.25 meters please note that CS values were only calculated per one quarter of a meter, and the precision is not 1 centimeter.

The jackknife tests indicated that the range of attraction did not change significantly when excluding any of the catch nights, experimental setups, or species groups. In model III the only significant aberration of the value of σ was when excluding the catches on 30.07.11, on this date the estimate changed to 14 with a 95% confidence interval 9.9-19.7, while all other results where within confidence levels (data not shown).

We notice that model I does not fit any of the data (Fig. D.5, top). Model II can only fit data with very different values for r (Fig. D.5, middle). While model III is able to fit both setup A and B using the same values for σ and r (Fig. D.5, bottom). Since Model III is the only model which can fit both experimental setups with the same values of r we have not included a comparison of models using information criteria.

Discussion

In this study we found a range of attraction for *Culicoides* at 15.25m. This means that the trap type used in this study should be separated by at least 30.5m (25.4-36.6) to enable independent sampling. When traps are placed closer than this, they will influence each other, competing for *Culicoides*. However, the range of attraction will also be extended when catch areas overlap, which is a novel result of this study. Thus, it is possible to cover a greater area from the same position by using more than one trap.

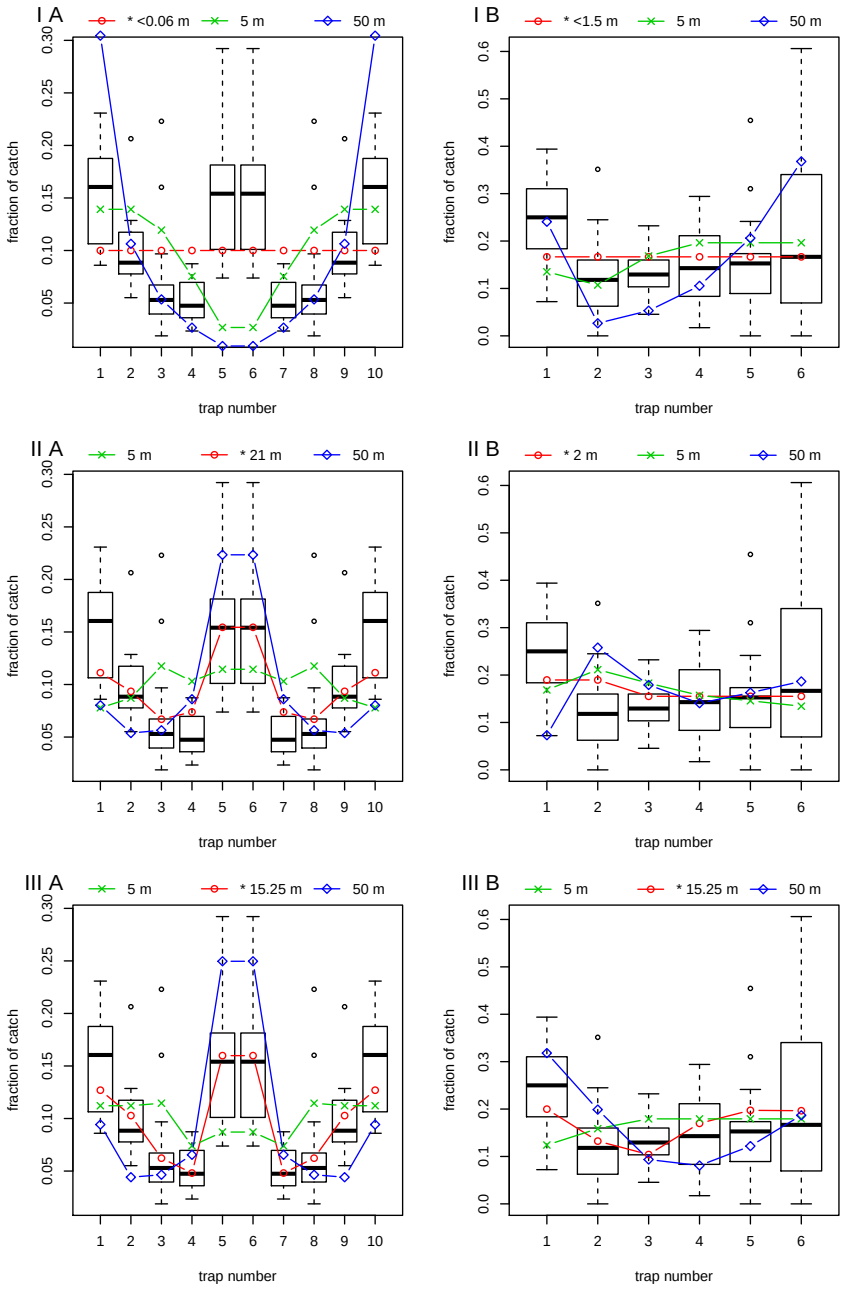

Figure D.5: Model fits to experimental setups. Predicted fractional catch per trap given model I (nearest trap), II (indirect light), and III (perceived light) (top, middle, bottom) with the field data as boxplots, for setup A and B (left, right). The prediction is the fractional part of the area covered within the range of attraction in Fig. D.2. The red line with circles is always the best fit of the model to the data. The predicted green and blue lines is fractional catch with a range of attraction of 5 or 50 meters. The ranges of attraction r^k giving the best fit for the k'th model were: $r^{IA/B} < 0.06/1.5$ m, $r^{IIA/B} = 21/ < 2$ m, and $r^{III} = 15.25$ m with $\sigma = 10$.

We used the relative levels of catches to estimate the range of attraction. This made the modelling independent of weather parameters causing changes in the abundance, e.g. wind speed or temperature. Spatial parameters, e.g. wind direction and location of hosts are likely to have an impact on the relative catches in the traps. However, we made an effort to compensate for this in the symmetrical shape of setup A, reversing the transects in setup B and by rotating setup B (Fig. D.1).

The range of attraction may differ between species. But Rigot *et al.* (2012) [108] found very similar ranges of attraction between vector species and vector species groups with overlapping confidence intervals. Because some species (e.g. *C. impunctatus*) may not be attracted by light as much as others [99], the range of attraction may be different for different species. Therefore, we conducted a jackknife test on the results by removing a species group one at a time, which showed that the results did not differ significantly when testing the species groups individually.

Trap efficiency is dependent on the background illumination, which can differ between sampling periods due to factors such as cloudiness, moon phase and time of sampling related to sunset (e.g. [113–117]). This is a potential source of bias and could result in different ranges of attraction between sampling periods. However, we tested this in a jackknife analysis, leaving out one catch night at a time, and found no significant difference in the estimate of the range of attraction. The background illumination is also more likely to impact on the σ parameter in model III because we expect that the *Culicoides* can distinguish light sources better under darker conditions. As previously stated, leaving out one of the catch nights did yield a significantly different estimate of σ (data not shown).

Model I (Nearest trap) failed to fit the data in experimental setup A and B. This model has a fixed range of attraction for each trap regardless of the distance to neighboring traps. This model type was used in the study of Bidlingmayer & Hem (1980) [114] to explain catch patterns of mosquitoes in traps without light, and it was recently used in another study to fit catches of *Culicoides* in light traps [108]. Given the physical properties of light, the effect of two neighboring light sources create an additive effect in the overlapping area, a main assumption in model II and model III. Thus, we can see from this study that the range of attraction from one point can be extended by using more traps, corresponding to a stronger light source.

Model III (Perceived light) was the only model to fit both experimental setup A and B (Fig. D.5). Thus, we regard 15.25m (12.7m-18.3m) as a reliable estimate of the range of attraction for one trap for *Culicoides* vectors. Rigot & Gilbert (2012) [108] found that the 8W Onderstepoort type traps had a range of

attraction of 29.6 (26.3m-31.9m). However, since the model used in that study failed to fit both experimental setups in our study, a more precise estimate may be obtained by using Model III from the present study. The fact that Venter *et al.* (2012) [109] found a range of attraction between 2 and 4 meters for the Onderstepoort type trap, could indicate that other unknown factors may be important when traps are allowed to catch the whole night through.

As stated in [118], the range of attraction covers three concepts: the distance at which specimens can physically reach the trap within a given time interval; the distance at which a specimen can detect the trap; and the distance at which a specimen shows directed movement towards the trap. If traps are allowed to sample a longer time, data can be influenced by other parameters such as wind direction and wind plumes created by host animals. If sampling time is too short, the specimens within the range of attraction may not be able to reach the trap before the sampling ends. To investigate the range of attraction and the influence of time, different sampling intervals would be needed, which is worth further research.

The distance from the *Culicoides* to a trap is also worth considering. We assumed general random flight with full attraction towards the traps within the range of attraction. However, the traps might attract a higher percentage of *Culicoides* in the vicinity closer to the traps compared to further out in the range of attraction, possibly proportional to light intensity.

In our models we assumed that *Culicoides* disperse evenly within the field. However, the abundance of *Culicoides* is likely to be higher near the cattle. We have tried to compensate for this by reversing the direction of the transects in setup B. Setup A compensates for this by the symmetry of the transect (Fig. D.1). Furthermore, the *Culicoides* may not be evenly dispersed when consecutive trapping is carried out because *Culicoides* within the range of attraction would be caught in the first trapping period and thus new *Culicoides* in the area would have to migrate in by random flight. To explain this pattern, a better fit may be obtained by fitting data to the circumference of the attraction area rather than the area itself. This could be worth investigating in future research. In both model II and model III, we assumed that all *Culicoides* caught in the traps approached the traps from the same height as the light sources, and therefore there were blind angles in the ends of the light tubes. If the *Culicoides* approached the traps from a lower height, the blind angle would be less pronounced. Although *Culicoides* have been caught at higher altitudes (e.g. [119]), and Venter *et al.* (2009) [120] caught most *Culicoides* at a height of 2.8m in South Africa, the main flying height for *Culicoides* vectors in northern Europe is still unknown.

In Models II and III the light distribution around each trap was anisotropic. We also simulated these two models using isotropic light, but for both models

anisotropic light fitted data better (data not shown). This indicates that the direction of the light tube in the trap is important, which has practical implications when catching *Culicoides*. If a certain area is to be monitored in a study, e.g. an enclosure with host animals, the catch size will depend on the angle of the trap to the area of interest. If trap catches are to be compared in a study, the standardization procedure should include direction of the light traps.

In this study, Model II modelled the light from each trap, resulting in higher light intensity when ranges of attraction overlap. This is comparable to a scent zone created isotropically around a host animal if there is no wind present. This model did not fit data as well as Model III did, where the *Culicoides* can perceive the individual sources of light at a distance and head for the strongest light source. This is an important biological finding and indicates that the *Culicoides* show directed movement towards a light source rather than a more random flight towards areas with higher light intensity. The implications of this finding is important for other studies using light trap catches to estimate the number of *Culicoides* (and possibly also other insects, e.g. mosquitoes) attracted to host animals. We have shown here that the vectors evaluate light sources at a distance. This behaviour is different from how we assume the vectors are attracted to host animals, i.e. following a plume of scent. Thus light trap catches may not represent the number and species of vectors attracted to hosts very well, and should be used with caution.

Conclusions

We tested three different models to fit two different field data sets, and showed that the *Culicoides* are likely to locate the light by evaluating the direction of the strongest light source in their field of view and then fly towards it rather than flying towards the nearest trap. We estimated the range of attraction for a single CDC 4W UV light trap to be 15.25m (12.7-18.3) perpendicular to the light tube. Therefore, we suggest that, in future studies, traps of this type are separated by at least 30.5m (25.4-36.6) in order to be independent. If they are placed closer than this, their interactions should be modelled as in model III in this study. Light traps may not represent the number of vectors attracted to hosts because the attraction behaviours are different.

Competing interest

The authors declare that they have no competing interests.

Author's contributions

This project is a part of the PhD project by Carsten Kirkeby at the Veterinary Institute at the Technical University of Denmark and the PhD project by Kaare Græsbøll at the Veterinary Institute and Department for Informatics and Mathematical Modelling, Technical University of Denmark. Carsten Kirkeby conceived the study, carried out the field work, participated in discussion of the modeling procedure and wrote the background, experimental setup, experimental setup results, discussion and conclusion sections of the manuscript and created fig. D.1. Kaare Græsbøll conceived the framework of model II and III, carried out all modeling and model fitting and wrote the modeling and the model results sections of this paper. Kaare also participated in the discussion and conclusions sections and created figs. D.2, D.3, D.4 & D.5. René Bødker participated in conceiving the study, planning of field experiments and discussion of the results. Anders Stockmarr participated in the planning of the field work and model fitting and evaluation. Lasse E. Christensen participated in conceiving and model III, the modeling procedure, model fitting and evaluation. All authors read and approved the final version of the manuscript.

Acknowledgements

Thanks to Peter and Lone Hald at Rubjerggaard for providing space for the field experiments, and Peter Lind at DTU Veterinary Institute for discussion of results and comments on the manuscript. This study was partially funded by EU grant FP7-261504 EDENext and is catalogued by the EDENext Steering Committee as EDENext070 (http://www.edenext.eu). The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

References

References are found sorted numerically from page 145 and alphabetically from 157.

Appendix E

Modelling spread of Bluetongue in Denmark: The code.

This PhD thesis does not include the actual code only the comments to the different modules and subroutine. The code was excluded because it is approximately 100 pages long.

A technical report (including full code) is available from DTU Orbit:

http://orbit.dtu.dk/fedora/objects/orbit:115176/datastreams/file_86c887d9-13fd-4622-a0e0-01c60aa3e61f/content

This appendix is an excerpt from this technical report, including all text except the code.

Introduction

This technical report was produced to make public the code produced as the main project of the PhD project by Kaare Græsbøll, with the title: "Modelling spread of Bluetongue and other vector borne diseases in Denmark and evaluation of intervention strategies".

The model

The main aim of the PhD thesis the code presented in this Technical report refer to, was to create a simulation model to predict the spread of Bluetongue and other vector borne diseases should they be introduced in Denmark. This model is presented in a publication with the title: "Simulating spread of Bluetongue Virus by flying vectors between hosts on pasture." [90] Which has been preliminary accepted for publication in Scientific Reports, and it is recommended to read this report in conjunction with the article. Mathematical expression and parameter values are not included in this report, but are to be found in the article.

The program code presented is for the case of Bluetongue Virus with cattle as hosts and biting midges as vectors. Bluetongue Virus (BTV) is a non-contagious infectious disease that infects ruminants. In Denmark the primary ruminant of concern is cattle in which the disease causes relative mild symptoms, but do reduce milk yield and increase the risk of spontaneous abortions. In certain breeds of sheep BTV can have a very high mortality, and therefore the disease is considered notifiable by the OIE. Bluetongue does non transmit directly between ruminants but requires to pass through a blood-sucking vector. Inside the vector BTV needs to replicate to a certain level and make it from the gut to saliva glands. This process is very sensitive to temperature and cannot be completed if the vector is at temperatures below 13°C. Therefore we take in meteorological data to calculate the extrinsic incubation period (EIP) a.k.a. the incubation time in the vectors.

What especially differentiates the model presented in this report with previous models on vector borne spread is that hosts can be distributed onto pasture areas. We wanted to to create a more process oriented approach so that the parameters describing spread of disease relates directly to parameters describing

flight patterns of vectors. For a more detailed description of the reasons to choose this particular model type see the PhD thesis.

The model share in-herd dynamics described in Gubbins (2008) [45] and Szmaragd (2009) [46] (Figure E.1), while the between herd model was presented in Græsbøll (2012) [90] (Figure E.2). Vaccination was modeled on UK data in a collaboration with the Pirbright Institute presented in Græsbøll (2012b) [121] using vaccination as described in Szmaragd (2010) [67].

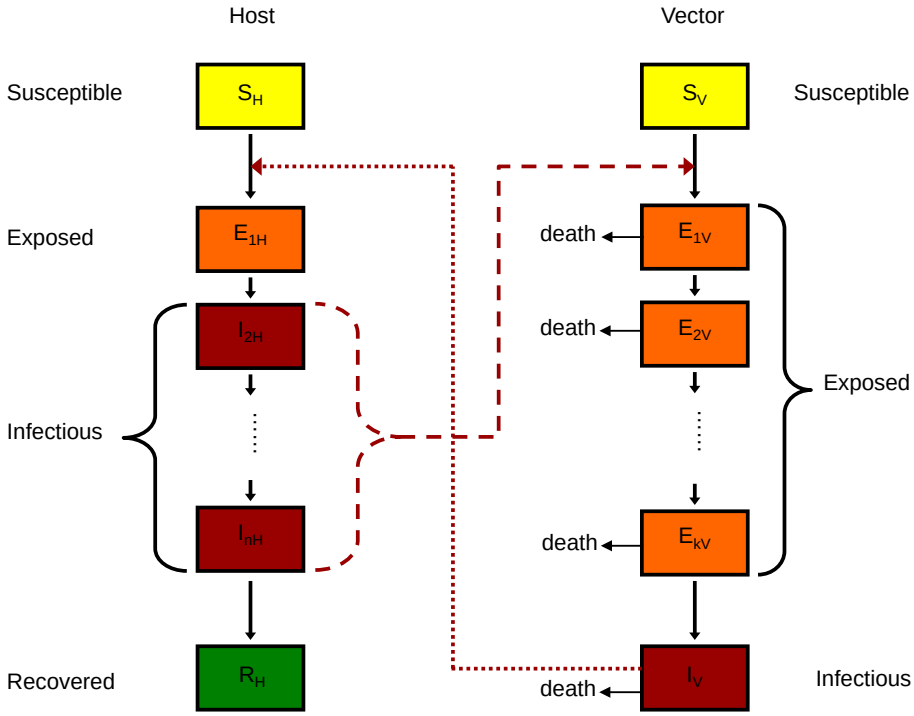

Figure E.1: The viraemia of the hosts is described by an extended Susceptible Exposed Infectious Recovered (SEIR) model (left), and the vectors are described by an extended SEI model (right). All movements between the stages in the model are governed by the probabilities listed in Græsbøll (2012) [90], in the code most parameters are defined in `gdata` (section E.2) or read in from the `input.txt` file (section E.6.1).

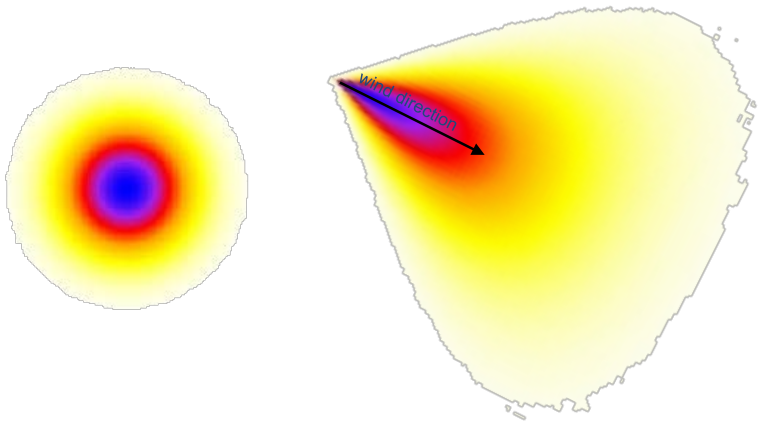

Figure E.2: Movement of vectors between herds are described either as active flight or passively carried by wind. This can be interpreted by 2D density distributions. (Left) probability density when vectors are moving by active flight to neighboring areas (arbitrary length scale, darker colors are equivalent to higher probability). (Right) probability density when vectors are carried by wind (arbitrary length scale). Mathematical expression can be found in Græsbøll (2012) [90].

Structure

The code was generated with an eye for speed, and alone for this reason the program was twice rebuild from scratch. The main issue with speed relates to the number of vectors: The number of vectors in high seasons is as many as 5,000 vectors per host. Given that there are approximately 1.7 million cattle in Denmark, this round up to almost 10 billion vectors, each need potentially be modeled for flight. This high number is never realistically reached, and optimizing for speed is therefore mainly a problem of best identifying relevant vectors to move. This is primarily achieved by keeping a list of all vectors and hosts with the disease, and only perform relevant operations: viraemia,

movement, death, etc. on these animals. In the program this is handled by two lists that cross-reference to each other and to the map. Using a list proved much faster (at least a factor of ten) than just having vectors and host located in a grid matching the map and scanning across areas looking for animals with disease.

The code is structured with a control file `dk.f90` which then call 13 other modules to do as explained in the following sections as listed below:

E.1	`dk.f90`	Determines the overall structure of the simulation.
E.2	`gdata.f90`	Defines the global data available to all modules.
E.3	`initialize.f90`	Initialises many of the variables and arrays.
E.4	`functions.f90`	Collates repeated functions.
E.5	`iohosts.f90`	Reads in pasture maps.
E.6	`iomodule.f90`	Reads input from the `input.txt` file.
E.7	`host.f90`	Handles host movements.
E.8	`temperatures.f90`	Handles temperature data.
E.9	`makemap.f90`	Distributes hosts onto pasture.
E.10	`viraemia.f90`	Emulates the viraemia of hosts and vectors.
E.11	`windy.f90`	Handles wind data.
E.12	`midge.f90`	Does the active and passive movements of vectors.
E.13	`bite.f90`	Transmission of disease between vectors and hosts.
E.14	`errorhandl.f90`	Various checks of the code.

Please note that section E.15 contains a full list of all modules and subroutines. Some parts of the code have been omitted to increase the ease of the read, these parts are mainly lines of codes that determines events when running in certain test modes of the program. However there are also omitted reallocation statements that handles sizes of arrays, these were omitted mostly due to space considerations. For a fully functional copy of the code please contact the author. The version number of the code presented is 9.7.4.

E.1 `dk.f90`

The main control file calls the subroutines that handles everything. The program is structured in this way because it makes it easy to turn subroutines on and off, which is very convenient in the test phase, and when testing different scenarios with regards to e.g. vaccinations or different countries. The code here presented contains two loops in `dk.f90`, one 'repeat' which is simple replicates of the simulations, and one 'day' that goes from day one to the designated 'end-day'. When running sensitivity analysis extra loops can be added to control

parameters outside these two loops. Furthermore the control file writes much of the output to file. Often it can be desirable to handle output in a separate module, I chose otherwise because the day-counter is in `dk.f90` it is very convenient to write output from this loop.

E.2 `gdata.f90`

As the program is build using the module functionality of Fortran 90 it is advantageous to define the global data arrays in a module which is then called by all other modules. Structuring data in this way ensures that this global data is permanently in the computer memory and only in one copy. Using a module to state global data also eliminates the need to move data between subroutines, or call it explicitly. This is very nice if much data is to be moved. However, this do comes on the expense of the detailed knowledge of what goes in and out of the specific subroutines. Thus it sometimes becomes tricky to keep track of all variable names, I sincerely recommend using the `implicit none`, and also keep a separate file which contains names of the most important variables and arrays and their structure. Here I present parts of my own "variable list", both as an example, but also to ease the read of the code. The number of parenthesis after a variable indicates the dimensionality of the array, so that none = scalar, () = vector, ()() = matrix, and so on. Variable inside [ ] indicates optional variables. the [rand] indicates that variables are only use when generating fields at random.

=== List of variables ==========

hosts(y)(x)(#C,C_id,[#S,S_id,]M_id) the spatial map linking to the list

csick_loc (C_id)(y,x,M_id,#H,SEIR) list of infected cattle

ssick_loc (S_id)(y,x,M_id,#H,SEIR) list of infected sheep

msick_loc(M_id)(y,x,C_id,#C,[S_id,#S,]#H,T_id,SEIR) list of infected midges

#C, #S, #M number of cattle, sheep, midges at a given location
#H = #C + #S total number of hosts at a given location

C_id, S_id, M_id the ID number to cross reference lists

=== gdata.f90 ===

−integer−

xsize	number of cells in x direction
ysize	number of cells in y direction
mstages	number of boxes used to emulate midge stages
ios	input/output error status
xstart	x value in cells of start of epidemic
ystart	y value in cells of start of epidemic
msick	count of locations of sick midges
n_farms	number of cells with hosts [rand]
nmax	maximum number of cells in a cluster to represent fields [rand]
nmean	mean number of cells in a cluster to represent fields [rand]
input_type	is data generated or read from file
area_sel	area select between predefined or own definition
infherds	count of infected herds
fly_c	count of flying events per time unit
day	current day of simulation
m_var	number of additional variables in the msick_loc array
mboxes	number of stages in the midge EIP
mbij	integer value of number of boxes to jump when emulate mboxes
h_var	number of additional variables in the csick_loc2 array
a_fly	the allocated size of the fly array
a_msick	the allocated size of the msick array
host_types	number of hosts types (cattle, sheep, etc)
fly()	array of fly to locations
msick_loc()	array with all infected midges
hstages()	number of hosts stages (no need to emulate <− temp indep.)
mth()	midges per host
hsick()	count of locations of sick hosts
n_hosts()	number of hosts
hosts()	array with hosts location
csick_loc()	locations of sick cattle
csick_loc_com()	history of infected cattle
ssick_loc()	locations of sick sheep
ssick_loc_com()	history of infected sheep
hsick_com()	counter for csick_loc_com+ssick_loc_com
a_hsick()	allocated size of hsick
a_hsick_c()	allocated size of hsick_loc_com
mk	parameter determining the precission of variables
pi	pi =3.1415....

−real−

loclen	local length scale for the random walk of midges
mrec	1 / EIP for midges

wile	width to length scale in wind spread cone
mmort	midge daily mortality
sizefac	cell side lenght in km
temperature	mean daily temperature
biterate	daily bite probability of midges
flytime	how long do midges stay in the wind [hours]
midgefly	fraction of midges that fly away from hosts
midgestay	fraction of midges that stay where no hosts
ave_dist	average distance between cells with hosts
meaninf	mean number of infected hosts per affected cell
vac_cov	fraction of hosts immune from vaccination
wind_s	wind speed in current cell
wind_a	angle of the wind in current cell
mbjr	real value of number of boxes to jump when emulate mboxes
hrec()	1/days for hosts to recover from infection
probtrans_mh()	probability of transmission midge to host
probtrans_hm()	probability of transmission host to midge
hosts_outside()	percentage hosts outside

`gdata.f90` also contains statements to turn host movement, movement restriction and a debugging mode (debugging statements have though mostly been removed from the displayed code) on and off.

Additional subroutines included are also primarily connected to the start-up of the program.

E.2.1 `init`

Defines many of the parameters characterizing which disease and the effect on specific hosts. These statements are though somewhat obsolete as they are also read in from the `input.txt` file (section E.6.1.1). But some are needed to initialize the size of certain data arrays.

E.2.2 `redistribute`

This subroutine reset all hosts and vectors arrays to the starting setup. This section is used for test runs when it is not desirable to resample all hosts onto pasture at every repetition.

E.2.3 `seeed`

The `seeed` subroutine ensures that the random seed is indeed random. (Trivia: Seeed is also a name of one of the author's favorite German bands.)

E.2.4 `alloc_cphid`

This subroutine allocate cph numbers to all pasture, so that hosts can be tracked back to their owners.

E.3 `initialize.f90`

The Initialize module in its full contains initialization of epidemic for both Denmark and the UK, and also the definition of the UK vaccination roll-out plan. Here I have only included the initialization of the Danish epidemic.

E.3.1 `init_inf`

It is possible to initiate using infected hosts or infected vectors. Here we also determine the area of first infection.

E.4 `functions.f90`

The `functions` module contains functions used throughout the program, but also subroutines to collect data and distribute flying vectors. From the original code some subroutines have been removed. The removed pieces of code did reallocation of arrays and error checks.

E.4.1 `BINV`

This function approximates a draw from a binomial distribution. Fortran 90 has only a call to a uniform random distribution. The simple solution would

be to do Bernoulli trials, however this method is significantly faster especially when the number of draws, n, is large and the probability, p, low. This combination of n and p is often fulfilled in the code given that there are often many vectors but transmission probability is very low. The algorithm was taken from Kachitvichyanukul and Schmeiser (1988) [122]. In this implementation the approximation crashes somewhere above $np = 500$ (probably due to rounding to zero), therefore for $np > 500$ the algorithm returns np. When running within reasonable disease parameter and vector abundance this cutoff is almost never in use.

E.4.2 `viraspread`

`viraspread` simply calculates the maximum distance virus has traveled from the starting point.

E.4.3 `herd_inf`

This subroutine reports the number of currently infected herds and the prevalence in the herds.

E.4.4 `flysort/flysortm`

When implementing dispersal of vectors, dispersing vectors was assigned to a temporary array so that the same vector was not able to fly more than once (see section E.12). The two modes (active and passive) of vector dispersal can run with the temporary array being either a list or a matrix mode depending on the expected number of flying vectors. The subroutines `flysort` and `flysortm` distributes vectors from the temporary array (list and matrix respectively) and makes sure that all indexing arrays are updated.

E.5 `iohosts.f90`

In most versions of the program pasture distribution is read in from pre-made files using the `read_mark_blocks` subroutine. These pre-maps are very large, and it is not feasible to store the number needed to ensure proper sampling.

However they do load much faster than generating new maps for every run, so for most runs sampling between pre-made maps was used (See fig. 2.3). The `o` part of `iohosts.f90` is represented by subroutines `find_herds` and `find_farms` designed to make maps of the spread.

E.5.1 `read_mark_blocks`

Reads in pasture data from files and distribute hosts onto these pasture areas. Pasture data for Denmark comes with a percentage of grass, therefore the program tries to distribute hosts onto pasture with high grass percentage first. It is also defined how many cattle per hectare that is maximum allowed.

E.5.2 `find_herds/find_farms`

Outputs infected herd/farm locations to file.

E.6 `iomodule.f90`

This module was prematurely named as an `io` module, considering that it only does input.

E.6.1 `read_inputfile`

To avoid numerous recompiles and to ease use for other users an `input.txt` was constructed which contains most of the parameters the program needs to run. `read_inputfile` reads and checks that input file. It also generate some parameters if the input file specified some random distribution. An example of the input file is included in section E.6.1.1.

E.6.1.1 input.txt

```
################### INPUT FILE for BTV-SPREAD program ####################
#      Author: Kaare Græsbøll (Graesboell) contact: kagr@imm.dtu.dk      #
#                  Version 3.4     28. Jul. 2012                        #
#                                                                       #
#                           Instructions:                               #
#                 -------------------------------                       #
#                                                                       #
#     A. SECTION:                                                       #
#     Choose Random distribution or CHR-Data                            #
#                                                                       #
#     B. SECTION:                                                       #
#     Choose region or set own values                                   #
#                                                                       #
#     C. SECTION:                                                       #
#     Change general parameters                                         #
#                                                                       #
#########################################################################
.
A. SECTION:
.
A1: Choose Random distribution (1) or CHR/CPH-Data (2):
.
Value.A1..........: 2
.
-------------------------------------------------------------------------
.
B. SECTION
.
B1: Choose region: Denmark (1), Lolland (2), Jutland (3), own definition (4),
.         UK farms (5), UK pasture sc 1,2,5 (8,9,10), DK-b sc 2,3,5 (11,12,13)
.
Value.B1..........: 12
.
   ---
   If A1=1 and B1=4 Define:
.
   side length of each grid cell (km): REAL
...sizefac........: 0.1
.
   Size of area (grids): INTEGER
...xsize..........: 5182
```

```
...ysize.........: 4880
.
   number of cattle: INTEGER !uk 67020
...farms+fields...: 67020
...cattle........: 6247964
.
   number of sheep: INTEGER !uk 58337, 19040196
...farms+fields...: 58337
...sheep.........: 21929519
.
.  Field distribution defined by binomial distribution: INTEGER
...mean.size.(np).: 1
...max.size.(n)...: 48
   ---
.
------------------------------------------------------------------------
.
C: SECTION
.
host types: (1) cattle (2) cattle + sheep: INTEGER (must be 2!)
.hosts_types......: 2
.
percentage of hosts outdoors: INTEGER
.hosts_out........: 0
.
stages of infection to mimic Gamma-dist for: INTEGER
.midges...........: 40
.cattle...........: 5
.sheep............: 7    !--- different because of dt
.
Recovery time for host (days): REAL
.cattle...........: 20.6
.sheep............: 16.4
.
width to length scale in wind spread cone: REAL
width.to.length...: 0.5
.
probability of transmission midge -> host: REAL
probtrans_mh......: 0.9
.
probability of transmission host -> midge: REAL
probtrans_hm......: 0.1
.
length of midge flight time (hours): REAL
```

```
flytime...........: 1.5
.
fraction of midges that fly with the wind: REAL
midgefly..........: 0.05
.
fraction of midges that stay where no hosts: REAL
midgestay.........: 0.05
.
local length scale for M random walk (km 50% of midges reach): REAL
loclen............: 0.5
.
number of midges per host (vectors per host): INTEGER
mth...............: 5000
.
vaccination cover (fraction): REAL
vac_cov...........: 0.
```

E.7 host.f90

This module contains the subroutine to handle host movements.

E.7.1 host_move

Host movements in the model are modeled by use of a transmission kernel. The kernel is fitted to data from "The Knowledgecenter of Agriculture, Dairy & Cattle farming" who tracks all movements of cattle in Denmark. Ideally a more process oriented network analysis should form the basis of movements in future models. However, using this simple kernel approach showed very little influence of cattle movement on the output of the model, so I do not believe that this approximation does much difference to the model.

E.8 temperatures.f90

This module handles temperature related subroutines. Specifically the interpolation from daily to hourly temperature and mapping temperature stations so that herds share temperatures with the spatially nearest data point. From the original code was deleted subroutines which handles temperatures for UK,

these where similar in function, and mostly differed because the structure of input data differs.

E.8.1 `temp_var_h`

Many of the parameters in this model are nonlinear with relation to temperature and most have cutoffs whereunder values go to zero. E.g. the Extrinsic Incubation Period (EIP) have a cutoff at $13.5°C$ so if the daily mean temperature is below this value virus cannot replicate inside vectors. However it is possible that a part of the day has temperatures above cutoff, and in these periods virus should be able to replicate. To emulate this we interpolate from daily minimum and maximum temperatures using a sinusoidal function to approximate hourly temperature data.

E.8.2 `read_temp_id`

The temperature data was provided by the Danish Meteorological Institute (www.dmi.dk) from 39 stations across Denmark. Data was processed for another project regarding Bluetongue (www.nordrisk.dk) where data was interpolated to a grid of 25 by 25 km. To increase the speed of the program a map of same resolution as the pasture map was made which indicate which temperature grid point the pasture should take temperature data from. This subroutine reads this pre-made map.

E.9 `makemap.f90`

This module contains the subroutine to make the grid cells (mark blocks) that represent the range vectors can locate hosts within.

E.9.1 `make_mark_blocks`

While the size of the grid cells in the simulation represents the vectors ability to locate hosts, these are also pasture areas. This subroutine collate the input data with regards to which pasture areas are owned by which farmers, and how much cattle are owned by the individual farmer. Cattle are then randomly distributed

onto pasture owned by the farmers according to criteria given in the `input.txt` file. (module approximately 90% finished.)

E.10 `viraemia.f90`

The `viraemia.f90` module deals with the viraemia of hosts and vectors.

E.10.1 `host_viraemia`

This sections moves the infected hosts through the stages in the SEIR model (figure 2.4).

E.10.2 `midge_viraemia`

This sections moves the exposed vectors through the exposed stages (figure 2.4). The code starts by handling the cases where all exposed and infectious vectors have moved away. Then runs through the list of exposed/infected midges updating the parameters according to temperature, and moving them through the states or killing them. In the end the code handles the case where all exposed and infectious vectors have died. I have left in some out-commented code that claims to ensure constant amount of midges, this is a remnant from when the code ran with a fixed amount of midges, so all midges having died were replaced in the susceptible stage. However, when number of midges became variable on a daily and seasonal basis, the interpretation of number of midges became equivalent to the number of susceptible midges. This was done because adjusting the total number of midges would be more computationally demanding than the out-commented line, and would also require further assumptions. An example: we sample daily between 1 and 5000 midges, day one there is 5000 midges, and we get 10 exposed midges, the next day we sample that there should be 1 midge. If this number is not assumed to be the susceptible number, we need to consider how to handle the 10 exposed midges from yesterday.

The observant reader might also have noted that the code seems to have 40 stages for the vectors, which is a lot different from the 11 as described in the theory section of Græsbøll (2012) [90]. From hereon the 40 partitions will be referred to as boxes to differentiate from the stages described in the theory. When this code was initially written the number of stages was only given as

an interval, and while e.g. Gubbins (2008) [45] handled this by assigning the number of stages for the EIP to each farm, this could not be achieved in this model because vectors move around, and how to handle a vector in stage four out of seven when it arrives at an area with 11 stages? This was solved by emulating stages using the 40 boxes. To exemplify: if there are ten stages at a given location (and for a given temperature) the vectors will jump four boxes forward each time we evaluate the viraemia. In this implementation we can think of the boxes as each representing 2.5% of the completion of EIP. This implementation allows for vectors to move between areas with different number of stages, and number of stages can also vary with temperature if needed. If the number of boxes divided by the number of stages is not an integer, the vectors will be distributed to the nearest boxes in proportions according to the decimal.

E.11 `windy.f90`

The `windy` module handles wind in the code.

E.11.1 `wind_data`

Wind is drawn from a random distribution. In the code there are some out-commented examples of how to generate anisotropic winds.

E.12 `midge.f90`

The module `midge` handles the movement of midges (vectors). Most importantly for the speed of the code, only vectors with the disease are allowed to move. Movement is designed to be handled in two different ways depending on the number of vectors to handle. When dealing with a small epidemic, every vector move is recorded into a list. When many vectors are flying all over the simulation box, the program can use a temporary matrix were all movements are subscribed to. The different modes are subsequently handled by the `flysort/flysortm` routines E.4.4, which places the flying vector in their landing areas. Parts of the code is out-commented this mostly different ways of handling vectors that flies outside the simulation box. The opted mode is that flights ending outside the simulation box is canceled. Out-commented modes are flights terminates at the edge, or flights outside box means instant dead.

E.12.1 midge_local

Local flight is simulated by assuming a Gaussian random walk. To generate coordinates the Box-Muller algorithm is utilized [123].

E.12.2 midge_wind

Midges carried by wind also have coordinates generated using the Box-Muller algorithm [123].

E.13 bite.f90

The `bite` module handles the transmission of virus between vectors and hosts. In the subroutines there are a lot of code which handles the different index arrays when bites leads to transmission. There is also different ways of handling vaccination included.

E.13.1 inf_hosts

In this subroutine hosts becomes exposed to the virus by bites from infectious midges.

E.13.2 inf_midges

In this subroutine susceptible vectors becomes exposed to the virus by biting infectious hosts. Notice that it is emulated that the titres of virus in blood is different in different stages of the disease, as has been seen in experimental data [124].

E.14 errorhandl.f90

In the end I show parts of the testing algorithm. When using index lists it becomes very important to make sure that these list cross-reference exactly so

that information is not lost or misinterpreted.

E.14.1 `findmidges`

Test if midge IDs are correctly handled in the host list.

E.14.2 `findhosts`

Test if host IDs are correctly handled in the host and vector lists.

E.15 Subroutine table

References – numerical

1. Velthuis, A. G. J., Saatkamp, H. W., Mourits, M. C. M., de Koeijer, A. A., and Elbers, A. R. W. Financial consequences of the dutch bluetongue serotype 8 epidemics of 2006 and 2007. *Preventive Veterinary Medicine* **93**(4), 294–304 (2010).

2. Purse, B. V. *et al.* Climate change and the recent emergence of bluetongue in Europe. *Nature Reviews Microbiology* **3**(2), 171–181 (2005).

3. Wilson, A. and Mellor, P. Bluetongue in Europe: vectors, epidemiology and climate change. *Parasitology Research* **103**, 69–77 (2008).

4. Hartemink, N. A. *et al.* Mapping the basic reproduction number (R_0) for vector-borne diseases: A case study on bluetongue virus. *Epidemics* **1**(3), 153 – 161 (2009).

5. Gould, E. A. and Higgs, S. Impact of climate change and other factors on emerging arbovirus diseases. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103**(2), 109 (2009).

6. Mellor, P. S. and Wittmann, E. J. Bluetongue virus in the mediterranean basin 1998 - 2001. *The Veterinary Journal* **164**(1), 20–37 (2002).

7. Alexander, K. A. *et al.* Evidence of natural bluetongue virus infection among african carnivores. *The American journal of tropical medicine and hygiene* **51**(5), 568 (1994).

8. Spreull, J. Malarial catarrhal fever (bluetongue) of sheep in South Africa. *J Comp Pathol Ther* **18**, 321–337 (1905).

9. Maclachlan, N. J. Bluetongue: History, global epidemiology, and pathogenesis. *Preventive Veterinary Medicine* **102**(2), 107–111 (2011).

10. Darpel, K. E. *et al.* Transplacental transmission of bluetongue virus 8 in cattle, UK. *Emerging infectious diseases* **15**(12), 2025 (2009).

11. Menzies, F. D. *et al.* Evidence for transplacental and contact transmission of bluetongue virus in cattle. *Vet. Rec.* **163**(7), 203–209 (2008).

12. Borkent, A. and Wirth, W. *World species of biting midges (Diptera: Ceratopogonidae).* American Museum of Natural History, (1997/2012).

13. Sanders, C. J. *et al.* Investigation of diel activity of Culicoides biting midges (diptera: Ceratopogonidae) in the united kingdom by using a vehicle-mounted trap. *Journal of medical entomology* **49**(3), 757–765 (2012).

14. Lassen, S., Nielsen, S., Skovgård, H., and Kristensen, M. Molecular identification of bloodmeals from biting midges (diptera: Ceratopogonidae: Culicoides latreille) in denmark. *Parasitology research* **108**(4), 823–829 (2011).

15. Lassen, S., Nielsen, S., and Kristensen, M. Identity and diversity of blood meal hosts of biting midges (diptera: Ceratopogonidae: Culicoides latreille) in denmark. *Parasites & Vectors* **5**, 143 (2012).

16. Mellor, P. S., Boorman, J., and Baylis, M. Culicoides biting midges: their role as arbovirus vectors. *Annual review of entomology* **45**(1), 307–340 (2000).

17. Carpenter, S. *et al.* Temperature dependence of the extrinsic incubation period of orbiviruses in Culicoides biting midges. *PLoS ONE* **6**(11), e27987 (2011).

18. Maan, S. *et al.* Novel bluetongue virus serotype from Kuwait. *Emerging infectious diseases* **17**(5), 886 (2011).

19. Paweska, J. T., Venter, G. J., and Mellor, P. S. Vector competence of south african Culicoides species for bluetongue virus serotype 1 (BTV-1) with special reference to the effect of temperature on the rate of virus replication in c. imicola and c. bolitinos. *Medical and Veterinary Entomology* **16**(1), 10–21 (2002).

20. Tabachnick, W. J. et al. Culicoides and the global epidemiology of bluetongue virus infection. *Veterinaria Italiana* **40**(3), 145 (2004).

21. Lorca-Oró, C. *et al.* Protection of spanish ibex (capra pyrenaica) against bluetongue virus serotypes 1 and 8 in a subclinical experimental infection. *PloS one* **7**(5), e36380 (2012).

22. Savini, G., MacLachlan, N. J., Sánchez-Vizcaino, J.-M., and Zientara, S. Vaccines against bluetongue in Europe. *Comparative Immunology, Microbiology and Infectious Diseases* **31**(2–3), 101–120 (2008).

23. Tabachnick, W. J. Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world. *Journal of Experimental Biology* **213**(6), 946–954 (2010).

24. Mellor, P. S., Carpenter, S., Harrup, L., Baylis, M., and Mertens, P. P. C. Bluetongue in Europe and the mediterranean basin: History of occurrence prior to 2006. *Preventive Veterinary Medicine* **87**(1–2), 4–20 (2008).

25. de Liberato, C. *et al.* Identification of Culicoides obsoletus (diptera: Ceratopogonidae) as a vector of bluetongue virus in central italy. *Veterinary Record* **156**(10), 301–304 (2005).

26. Savini, G. *et al.* Bluetongue virus isolations from midges belonging to the Obsoletus complex (Culicoides, diptera: Ceratopogonidae) in italy. *Veterinary Record* **157**(5), 133–139 (2005).

27. Jennings, D. M. and Mellor, P. S. The vector potential of British Culicoides species for bluetongue virus. *Veterinary Microbiology* **17**(1), 1 – 10 (1988).

28. Mellor, P. S. Culicoides as potential orbivirus vectors in Europe. In *Bluetongue, African horse sickness and related orbiviruses (TE Walton & BI Osburn, eds). Proc. Second International Symposium, Paris*, 17–21, (1991).

29. Mellor, P. S. Infection of the vectors and bluetongue epidemiology in Europe. *Veterinaria Italiana* **40**(3), 167 (2004).

30. Carpenter, S., Lunt, H. L., Arav, D., Venter, G. J., and Mellor, P. S. Oral susceptibility to bluetongue virus of Culicoides (Diptera: Ceratopogonidae) from the United Kingdom. *J. Med. Entomol.* **43**, 73–78 (2006).

31. Kampen, H. and Werner, D. Three years of bluetongue disease in central Europe with special reference to germany: what lessons can be learned? *Wiener Klinische Wochenschrift* **122**, 31–39 (2010).

32. Stott, J., Oberst, R., Channell, M., and Osburn, B. Genome segment reassortment between two serotypes of bluetongue virus in a natural host. *Journal of virology* **61**(9), 2670–2674 (1987).

33. Hoffmann, B. *et al.* Novel orthobunyavirus in Cattle, Europe, 2011. *Emerging Infect. Dis.* **18**(3), 469–472 (2012).

34. Rasmussen, L. D. *et al.* Culicoides as vectors of schmallenberg virus [letter]. *Emerging Infectious Diseases* (2012).

35. Gibbens, N. Schmallenberg virus: a novel viral disease in northern Europe. *Vet. Rec.* **170**, 58 (2012).

36. EFSA. "Schmallenberg" virus: likely epidemiological scenarios and data needs. Technical Report 2012:EN-241, European Food Safety Authority (EFSA), Parma, Italy, (2012).

37. Charles, J. Akabane virus. *veterinary clinics of North America food animal practice* **10**, 525–525 (1994).

38. Pinheiro, F. P. *et al.* Oropouche virus, a review of clinical, epidemiological and ecological findings. *Am J Trop Med Hyg* **30**, 149–160 (1981).

39. Yadin, H. *et al.* Epizootic haemorrhagic disease virus type 7 infection in cattle in Israel. *Veterinary Record* **162**(2), 53–56 (2008).

40. Yanase, T. *et al.* The resurgence of shamonda virus, an african simbu group virus of the genus orthobunyavirus, in japan. *Archives of virology* **150**(2), 361–369 (2005).

41. Rodriguez, M., Hooghuis, H., and Castaño, M. African horse sickness in Spain. *Veterinary Microbiology* **33**(1-4), 129–142 (1992).

42. Mellor, P. and Hamblin, C. African horse sickness. *Veterinary research* **35**(4), 445–466 (2004).

43. Diekmann, O. and Heesterbeek, J. A. P. *Mathematical epidemiology of infectious diseases: model building, analysis, and interpretation*, volume 5. Wiley, (2000).

44. Hartemink, N. Vector-borne diseases: the basic reproduction number R_0 and risk maps. http://igitur-archive.library.uu.nl/dissertations/20091126-200122/hartemink.pdf, (2009).

45. Gubbins, S., Carpenter, S., Baylis, M., Wood, J. L. N., and Mellor, P. S. Assessing the risk of bluetongue to uk livestock: uncertainty and sensitivity analyses of a temperature-dependent model for the basic reproduction number. *Journal of The Royal Society Interface* **5**(20), 363–371 (2008).

46. Szmaragd, C. *et al.* A modeling framework to describe the transmission of bluetongue virus within and between farms in great britain. *PLoS ONE* **4**(11), e7741 (2009).

47. de Koeijer, A. A. *et al.* Quantitative analysis of transmission parameters for bluetongue virus serotype 8 in western Europe in 2006. *Veterinary research* **42**(1), 1–9 (2011).

48. Pioz, M. *et al.* Why did bluetongue spread the way it did? environmental factors influencing the velocity of bluetongue virus serotype 8 epizootic wave in france. *PloS one* **7**(8), e43360 (2012).

49. Boender, G. J. *et al.* Quantitative analysis of spatial transmission for bluetongue virus serotype 8 in western Europe in 2006 and 2007. http://sydney.edu.au/vetscience/research/geovet/documents/presentations/Boender_49.pdf, (2010). GEOVET Conference, Sydney.

50. Turner, J., Bowers, R. G., and Baylis, M. Modelling bluetongue virus transmission between farms using animal and vector movements. *Sci. Rep.* **2** (2012).

51. Sellers, R. and Pedgley, D. Possible windborne spread to western Turkey of bluetongue virus in 1977 and of akabane virus in 1979. *Journal of hygiene* **95**(1), 149–158 (1985).

52. Braverman, Y. and Chechik, F. Air streams and the introduction of animal diseases borne on Culicoides (diptera, ceratopogonidae) into israel. *Revue scientifique et technique (International Office of Epizootics)* **15**(3), 1037 (1996).

53. Alba, A., Casal, J., and Domingo, M. Possible introduction of bluetongue into the Balearic islands, Spain, in 2000, via air streams. *Veterinary record* **155**(15), 460–461 (2004).

54. Ducheyne, E. *et al.* Quantifying the wind dispersal of Culicoides species in Greece and Bulgaria. *Geospatial Health* **1**(2), 177–189 (2007).

55. Hendrickx, G. *et al.* A wind density model to quantify the airborne spread of Culicoides species during north-western Europe bluetongue epidemic, 2006. *Preventive Veterinary Medicine* **87**(1-2), 162–181 (2008).

56. Jones, A., Thomson, D., Hort, M., and Devenish, B. The U.K. met office's next-generation atmospheric dispersion model, NAME III. In Air Pollution Modeling and Its Application XVII, Borrego, C. and Norman, A.-L., editors, 580–589. Springer US (2007).

57. Gloster, J., Mellor, P., Burgin, L., Sanders, C., and Carpenter, S. Will bluetongue come on the wind to the United Kingdom in 2007? *Veterinary Record* **160**(13), 422–426 (2007).

58. Burgin, L. E. *et al.* Investigating incursions of bluetongue virus using a model of long-distance Culicoides biting midge dispersal. *Transboundary and Emerging Diseases* (2012).

59. Sedda, L. *et al.* A new algorithm quantifies the roles of wind and midge flight activity in the bluetongue epizootic in northwest Europe. *Proc. Biol. Sci.* **279**(1737), 2354–2362 (2012).

60. Bladt, M. A review on phase-type distributions and their use in risk theory. *Astin Bulletin* **35**(1), 145–161 (2005).

61. Gubbins, S. *et al.* Scaling from challenge experiments to the field: Quantifying the impact of vaccination on the transmission of bluetongue virus serotype 8. *Preventive Veterinary Medicine* **105** (2012).

62. Savini, G. *et al.* Assessment of efficacy of a bivalent btv-2 and btv-4 inactivated vaccine by vaccination and challenge in cattle. *Veterinary microbiology* **133**(1), 1–8 (2009).

63. Viswanathan, G. M. *et al.* Optimizing the success of random searches. *Nature* **401**(6756), 911–914 (1999).

64. Kirkeby, C., Bødker, R., Stockmarr, A., Lind, P., and Heegaard, P. M. Quantifying dispersal of european culicoides (diptera: Ceratopogonidae) vectors between farms using a novel mark-release-recapture technique. *PloS one* **8**(4), e61269 (2013).

65. Nielsen, S., Nielsen, B. O., Axelsen, J. A., and Fotel, F. Forurening af egeløkke lung og opformering af mitter. Technical report, Miljøstyrelsen, (1996).

66. Sanders, C. J. *et al.* Influence of season and meteorological parameters on flight activity of Culicoides biting midges. *Journal of Applied Ecology* **48**(6), 1355–1364 (2011).

67. Szmaragd, C. *et al.* The spread of bluetongue virus serotype 8 in great britain and its control by vaccination. *PloS one* **5**(2), e9353 (2010).

68. Carpenter, S., Mellor, P., and Torr, S. Control techniques for Culicoides biting midges and their application in the UK and northwestern palaearctic. *Medical and veterinary entomology* **22**(3), 175–187 (2008).

69. Barnard, B. Some factors governing the entry of Culicoides spp.(diptera: Ceratopogonidae) into stables. *Onderstepoort Journal of Veterinary Research* **64**, 227–233 (1997).

70. Silbermayr, K., Hackländer, K., Doscher, C., Koefer, J., and Fuchs, K. A spatial assessment of Culicoides spp. distribution and bluetongue disease risk areas in Austria. *Berliner und Münchener tierärztliche Wochenschrift* **124**(5-6), 228 (2011).

71. Mullens, B. A., Gerry, A. C., Lysyk, T. J., and Schmidtmann, E. T. Environmental effects on vector competence and virogenesis of bluetongue virus in Culicoides: interpreting laboratory data in a field context. *Vet. Ital.* **40**, 160–166 (2004).

72. Gerry, A. C. and Mullens, B. A. Seasonal abundance and survivorship of Culicoides sonorensis (diptera: Ceratopogonidae) at a southern california dairy, with reference to potential bluetongue virus transmission and persistence. *Entomological Society of America* **37**, 675–688 (2000).

73. Melville, L. *et al.* Characteristics of naturally-occurring bluetongue viral infections of cattle. In Bluetongue Disease in Southeast Asia and the Pacific., St George, T. and Peng, K., editors, Canberra: ACIAR proceedings series, 245–250, (1996).

74. Baylis, M. *et al.* Evaluation of housing as a means to protect cattle from Culicoides biting midges, the vectors of bluetongue virus. *Medical and veterinary entomology* **24**(1), 38–45 (2010).

75. Calvete, C. *et al.* Entry of bluetongue vector Culicoides imicola into livestock premises in Spain. *Medical and veterinary entomology* **23**(3), 202–208 (2009).

76. Meiswinkel, R., Baylis, M., and Labuschagne, K. Stabling and the protection of horses from Culicoides bolitinos (diptera: Ceratopogonidae), a recently identified vector of african horse sickness. *Bulletin of entomological research* **90**(06), 509–515 (2000).

77. Dufour, B., Moutou, F., Hattenberger, A. M., and Rodhain, F. Global change: impact, management, risk approach and health measures–the case of Europe. *Rev. - Off. Int. Epizoot.* **27**, 529–550 (2008).

78. Wilson, A. J. and Mellor, P. S. Bluetongue in Europe: past, present and future. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **364**, 2669–2681 (2009).

79. Guis, H. *et al.* Use of high spatial resolution satellite imagery to characterize landscapes at risk for bluetongue. *Vet. Res.* **38**, 669–683 (2007).

80. Danish ministry of food, agriculture, and fisheries. www.uk.foedevarestyrelsen.dk/AnimalHealth/Bluetongue/forside.htm, (2012).

81. Kirkeby, C., Bødker, R., Stockmarr, A., and Enøe, C. Association between land cover and Culicoides (diptera: Ceratopogonidae) breeding sites on four danish cattlefarms. *Entomologica Fennica* **20**(4), 228–232 (2009).

82. Ninio, C., Augot, D., Dufour, B., and Depaquit, J. Emergence of Culicoides obsoletus from indoor and outdoor breeding sites. *Veterinary Parasitology* **183**, 125–129 (2011).

83. Nielsen, S., Nielsen, B. O., Axelsen, J. A., and Fotel, F. Aktivitet af mitter på græsningsarealer ved egeløkke lung. Technical report, Miljøstyrelsen, (2003).

84. Ortega, M. D., Mellor, P. S., Rawlings, P., and Pro, M. J. The seasonal and geographical distribution of Culicoides imicola, C. pulicaris group and C. obsoletus group biting midges in central and southern Spain. *Arch. Virol. Suppl.* **14**, 85–91 (1998).

85. Baylis, M., O'Connell, L., and Mellor, P. S. Rates of bluetongue virus transmission between Culicoides sonorensis and sheep. *Med. Vet. Entomol.* **22**, 228–237 (2008).

86. Carpenter, S. *et al.* Experimental infection studies of UK Culicoides species midges with bluetongue virus serotypes 8 and 9. *Vet. Rec.* **163**, 589–592 (2008).

87. Keeling, M. J. Models of foot-and-mouth disease. *Proc. Biol. Sci.* **272**(1569), 1195–1202 (2005).

88. Menzies, F. D. *et al.* Risk assessment, targeted surveillance and policy formulation: practical experiences with bluetongue. In *International Conference on Animal Health Surveillance (ICAHS), Lyon, France, 17-20 May, 2011.*, number 59/60, 113–115. Association pour l'Étude de l'Épidémiologie des Maladies Animales, (2011).

89. Marangon, S., Cecchinato, M., and Capua, I. Use of vaccination in avian influenza control and eradication. *Zoonoses and Public Health* **55**(1), 65–72 (2008).

90. Græsbøll, K., Bødker, R., Enøe, C., and Christiansen, L. E. Simulating spread of bluetongue virus by flying vectors between hosts on pasture. *Sci. Rep.* **2**(863) (2012).

91. Jewell, C. P., Keeling, M. J., and Roberts, G. O. Predicting undetected infections during the 2007 foot-and-mouth disease outbreak. *Journal of The Royal Society Interface* **6**(41), 1145–1151 (2009).

92. UK meteorological office. met office integrated data archive system (MIDAS) land and marine surface stations data (1853-current), [internet].NCAS british atmospheric data centre, 2012. http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_ukmomidas, (2011).

93. Department for environment food and rural affairs. Report on the distribution of bluetongue infection in Great Britain. Defra, London, U.K. 26pp., (2008).

94. Boender, G. J. *et al.* Risk maps for the spread of highly pathogenic avian influenza in poultry. *PLoS Computational Biology* **3**(4), e71 (2007).

95. Dyce, A. L. The recognition of nulliparous and parous Culicoides (diptera: Ceratopogonidae) without dissection. *Australian Journal of Entomology* **8**(1), 11–15 (1969).

96. Clinger, W. and Van Ness, J. W. On unequally spaced time points in time series. *The Annals of Statistics* **4**(4), 736–745 (1976).

97. Kirkpatrick, S., Gelatt Jr, C. D., and Vecchi, M. P. Optimization by simulated annealing. *Science* **220**(4598), 671–680 (1983).

98. Aarts, E. H. L. and Laarhoven, P. J. M. Simulated annealing: an introduction. *Statistica Neerlandica* **43**(1), 31–52 (1989).

99. Holmes, P. R. and Birley, M. H. An improved method for survival rate analysis from time series of haematophagous dipteran populations. *The Journal of Animal Ecology* **56**(2), 427–440 (1987).

100. Venter, G. J. *et al.* Comparison of the efficiency of five suction light traps under field conditions in south africa for the collection of Culicoides species. *Veterinary parasitology* **166**(3), 299–307 (2009).

101. Allan, S. A., Day, J. F., and Edman, J. D. Visual ecology of biting flies. *Annual review of entomology* **32**(1), 297–314 (1987).

102. Purse, B. V. *et al.* Modelling the distributions of Culicoides bluetongue virus vectors in sicily in relation to satellite-derived climate variables. *Medical and Veterinary Entomology* **18**(2), 90–101 (2004).

103. Conte, A., Goffredo, M., Ippoliti, C., and Meiswinkel, R. Influence of biotic and abiotic factors on the distribution and abundance of Culicoides imicola and the obsoletus complex in italy. *Veterinary parasitology* **150**(4), 333–344 (2007).

104. Takken, W. *et al.* The phenology and population dynamics of Culicoides spp. in different ecosystems in the netherlands. *Preventive veterinary medicine* **87**(1), 41–54 (2008).

105. Odetoyinbo, J. A. Preliminary investigation on the use of a light-trap for sampling malaria vectors in the gambia. *Bulletin of the World Health Organization* **40**(4), 547 (1969).

106. Baker, R. R. and Sadovy, Y. The distance and nature of the light-trap response of moths. *Nature* **276**, 818–821 (1978).

107. Truxa, C. and Fiedler, K. Attraction to light–from how far do moths (lepidoptera) return to weak artificial sources of light? *Eur. J. Entomol* **109**(1), 77–84 (2012).

108. Rigot, T. and Gilbert, M. Quantifying the spatial dependence of Culicoides midge samples collected by onderstepoort-type blacklight traps: an experimental approach to infer the range of attraction of light traps. *Medical and Veterinary Entomology* **26**(2), 152–161 (2012).

109. Venter, G. J., Majatladi, D. M., Labuschagne, K., Boikanyo, S. N. B., and Morey, L. The attraction range of the Onderstepoort 220v light trap for Culicoides biting midges as determined under south african field conditions. *Veterinary Parasitology* **in press** (2012).

110. Campbell, J. A. and Pelham-Clinton, E. X.a taxonomic review of the british species of culicoides latreille (diptera, ceratopogonidæ). *Proceedings of the Royal Society of Edinburgh. Section B. Biology* **67**(03), 181–302 (1960).

111. Little, R. J. A. and Rubin, D. B. *Statistical analysis with missing data*, volume 1. Wiley New York, (1987).

112. Madsen, H. *Time series analysis*. CRC Press, (2008).

113. Bowden, J. and Church, B. The influence of moonlight on catches of insects in light-traps in africa. part ii. the effect of moon phase on light-trap catches. *Bulletin of Entomological Research* **63**(01), 129–142 (1973).

114. Bidlingmayer, W. and Hem, D. The range of visual attraction and the effect of competitive visual attractants upon mosquito (diptera: Culicidae) flight. *Bulletin of Entomological Research* **70**(02), 321–342 (1980).

115. Muirhead-Thomson, R. et al. *Trap responses of flying insects. The influence of trap design on capture efficiency.* Academic Press Limited, (1991).

116. Nowinszky, L. Nocturnal illumination and night flying insects. *Appl. Ecol. Environ. Res* **2**, 17–52 (2004).

117. Silver, J. *Mosquito ecology: field sampling methods*. Springer Verlag, (2007).

118. Wall, C. and Perry, J. Range of action of moth sex-attractant sources. *Entomologia experimentalis et applicata* **44**(1), 5–14 (1987).

119. Chapman, J., Reynolds, D., Smith, A., Smith, E., and Woiwod, I. An aerial netting study of insects migrating at high altitude over England. *Bulletin of Entomological Research, London* **94**(2), 123–136 (2004).

120. Venter, G., Hermanides, K., Boikanyo, S., Majatladi, D., and Morey, L. The effect of light trap height on the numbers of Culicoides midges collected under field conditions in south africa. *Veterinary parasitology* **166**(3), 343–345 (2009).

121. Græsbøll, K., Sumner, T., Enøe, C., Christiansen, L. E., and Gubbins, S. A comparison of dynamics in two models for the spread of a vector borne disease. *Sci. Rep.* **?**(Submitted) (2012/2013).

122. Kachitvichyanukul, V. and Schmeiser, B. W. Binomial random variate generation. *Communications of the ACM* **31**(2), 216–222 (1988).

123. Box, G. E. P. and Muller, M. E. A note on the generation of random normal deviates. *The Annals of Mathematical Statistics* **29**(2), 610–611 (1958).

124. Barratt-Boyes, S. M. and MacLachlan, N. J. Dynamics of viral spread in bluetongue virus infected calves. *Veterinary microbiology* **40**(3), 361–371 (1994).

References – alphabetical

Aarts, E. H. L., and Laarhoven, P. J. M. Simulated annealing: an introduction. *Statistica Neerlandica 43*, 1 (1989), 31–52.

Alba, A., Casal, J., and Domingo, M. Possible introduction of bluetongue into the Balearic islands, Spain, in 2000, via air streams. *Veterinary record 155*, 15 (2004), 460–461.

Alexander, K. A., MacLachlan, N. J., Kat, P. W., House, C., O'Brien, S. J., Lerche, N. W., Sawyer, M., Frank, L. G., Holekamp, K., Smale, L., et al. Evidence of natural bluetongue virus infection among african carnivores. *The American journal of tropical medicine and hygiene 51*, 5 (1994), 568.

Allan, S. A., Day, J. F., and Edman, J. D. Visual ecology of biting flies. *Annual review of entomology 32*, 1 (1987), 297–314.

Baker, R. R., and Sadovy, Y. The distance and nature of the light-trap response of moths. *Nature 276* (1978), 818–821.

Barnard, B. Some factors governing the entry of Culicoides spp.(diptera: Ceratopogonidae) into stables. *Onderstepoort Journal of Veterinary Research 64* (1997), 227–233.

Barratt-Boyes, S. M., and MacLachlan, N. J. Dynamics of viral spread in bluetongue virus infected calves. *Veterinary microbiology 40*, 3 (1994), 361–371.

Baylis, M., O'Connell, L., and Mellor, P. S. Rates of bluetongue virus transmission between Culicoides sonorensis and sheep. *Med. Vet. Entomol. 22* (Sep 2008), 228–237.

Baylis, M., Parkin, H., Kreppel, K., Carpenter, S., Mellor, P., and McIntyre, K. Evaluation of housing as a means to protect cattle from Culicoides biting midges, the vectors of bluetongue virus. *Medical and veterinary entomology 24*, 1 (2010), 38–45.

Bidlingmayer, W., and Hem, D. The range of visual attraction and the effect of competitive visual attractants upon mosquito (diptera: Culicidae) flight. *Bulletin of Entomological Research 70*, 02 (1980), 321–342.

Bladt, M. A review on phase-type distributions and their use in risk theory. *Astin Bulletin 35*, 1 (2005), 145–161.

Boender, G. J., de Koeijer, A. A., Staubach, C., Gethmann, J. M., Meroc, E., Nodelijk, H. A., and Elbers, A. R. W. Quantitative analysis of spatial transmission for bluetongue virus serotype 8 in western Europe in 2006 and 2007. http://sydney.edu.au/vetscience/research/geovet/documents/presentations/Boender_49.pdf, 2010. GEOVET Conference, Sydney.

Boender, G. J., Hagenaars, T. J., Bouma, A., Nodelijk, G., Elbers, A. R. W., de Jong, M. C. M., and van Boven, M. Risk maps for the spread of highly pathogenic avian influenza in poultry. *PLoS Computational Biology 3*, 4 (2007), e71.

Borkent, A., and Wirth, W. *World species of biting midges (Diptera: Ceratopogonidae).* American Museum of Natural History, 1997/2012.

Bowden, J., and Church, B. The influence of moonlight on catches of insects in light-traps in africa. part ii. the effect of moon phase on light-trap catches. *Bulletin of Entomological Research 63*, 01 (1973), 129–142.

Box, G. E. P., and Muller, M. E. A note on the generation of random normal deviates. *The Annals of Mathematical Statistics 29*, 2 (1958), 610–611.

Braverman, Y., and Chechik, F. Air streams and the introduction of animal diseases borne on Culicoides (diptera, ceratopogonidae) into israel. *Revue scientifique et technique (International Office of Epizootics) 15*, 3 (1996), 1037.

Burgin, L. E., Gloster, J., Sanders, C., Mellor, P. S., Gubbins, S., and Carpenter, S. Investigating incursions of bluetongue virus using a model of long-distance Culicoides biting midge dispersal. *Transboundary and Emerging*

Diseases (2012).

Calvete, C., Estrada, R., Miranda, M., Del Rio, R., Borrás, D., Beldron, F., Martínez, A., Calvo, A., and Lucientes, J. Entry of bluetongue vector Culicoides imicola into livestock premises in Spain. *Medical and veterinary entomology 23*, 3 (2009), 202–208.

Campbell, J. A., and Pelham-Clinton, E. X.a taxonomic review of the british species of culicoides latreille (diptera, ceratopogonidæ). *Proceedings of the Royal Society of Edinburgh. Section B. Biology 67*, 03 (1960), 181–302.

Carpenter, S., Lunt, H. L., Arav, D., Venter, G. J., and Mellor, P. S. Oral susceptibility to bluetongue virus of Culicoides (Diptera: Ceratopogonidae) from the United Kingdom. *J. Med. Entomol. 43* (Jan 2006), 73–78.

Carpenter, S., McArthur, C., Selby, R., Ward, R., Nolan, D. V., Luntz, A. J., Dallas, J. F., Tripet, F., and Mellor, P. S. Experimental infection studies of UK Culicoides species midges with bluetongue virus serotypes 8 and 9. *Vet. Rec. 163* (Nov 2008), 589–592.

Carpenter, S., Mellor, P., and Torr, S. Control techniques for Culicoides biting midges and their application in the UK and northwestern palaearctic. *Medical and veterinary entomology 22*, 3 (2008), 175–187.

Carpenter, S., Wilson, A., Barber, J., Veronesi, E., Mellor, P., Venter, G., and Gubbins, S. Temperature dependence of the extrinsic incubation period of orbiviruses in Culicoides biting midges. *PLoS ONE 6*, 11 (11 2011), e27987.

Chapman, J., Reynolds, D., Smith, A., Smith, E., and Woiwod, I. An aerial netting study of insects migrating at high altitude over England. *Bulletin of Entomological Research, London 94*, 2 (2004), 123–136.

Charles, J. Akabane virus. *veterinary clinics of North America food animal practice 10* (1994), 525–525.

Clinger, W., and Van Ness, J. W. On unequally spaced time points in time series. *The Annals of Statistics 4*, 4 (1976), 736–745.

Conte, A., Goffredo, M., Ippoliti, C., and Meiswinkel, R. Influence of biotic and abiotic factors on the distribution and abundance of Culicoides imicola and the obsoletus complex in italy. *Veterinary parasitology 150*, 4 (2007), 333–344.

Darpel, K. E., Batten, C. A., Veronesi, E., Williamson, S., Anderson, P., Dennison, M., Clifford, S., Smith, C., Philips, L., Bidewell, C., et al. Transplacental transmission of bluetongue virus 8 in cattle, UK. *Emerging infectious diseases 15*, 12 (2009), 2025.

de Koeijer, A. A., Boender, G. J., Nodelijk, G., Staubach, C., Meroc, E., and Elbers, A. R. W. Quantitative analysis of transmission parameters for bluetongue virus serotype 8 in western Europe in 2006. *Veterinary research 42*, 1 (2011), 1–9.

de Liberato, C., Scavia, G., Lorenzetti, R., Scaramozzino, P., Amaddeo, D., Cardeti, G., Scicluna, M., Ferrari, G., and Autorino, G. L. Identification of Culicoides obsoletus (diptera: Ceratopogonidae) as a vector of bluetongue virus in central italy. *Veterinary Record 156*, 10 (2005), 301–304.

Diekmann, O., and Heesterbeek, J. A. P. *Mathematical epidemiology of infectious diseases: model building, analysis, and interpretation*, vol. 5. Wiley, 2000.

Ducheyne, E., De Deken, R., Bécu, S., Codina, B., Nomikou, K., Mangana-Vougiaki, O., Georgiev, G., Purse, B., and Hendrickx, G. Quantifying the wind dispersal of Culicoides species in Greece and Bulgaria. *Geospatial Health 1*, 2 (2007), 177–189.

Dufour, B., Moutou, F., Hattenberger, A. M., and Rodhain, F. Global change: impact, management, risk approach and health measures–the case of Europe. *Rev. - Off. Int. Epizoot. 27* (Aug 2008), 529–550.

Dyce, A. L. The recognition of nulliparous and parous Culicoides (diptera: Ceratopogonidae) without dissection. *Australian Journal of Entomology 8*, 1 (1969), 11–15.

EFSA. "Schmallenberg" virus: likely epidemiological scenarios and data needs. Technical Report 2012:EN-241, European Food Safety Authority (EFSA), Parma, Italy, February 2012.

Gerry, A. C., and Mullens, B. A. Seasonal abundance and survivorship of Culicoides sonorensis (diptera: Ceratopogonidae) at a southern california dairy, with reference to potential bluetongue virus transmission and persistence. *Entomological Society of America 37* (2000), 675–688.

Gibbens, N. Schmallenberg virus: a novel viral disease in northern Europe. *Vet. Rec. 170* (Jan 2012), 58.

Gloster, J., Mellor, P., Burgin, L., Sanders, C., and Carpenter, S. Will bluetongue come on the wind to the United Kingdom in 2007? *Veterinary Record 160*, 13 (2007), 422–426.

Gould, E. A., and Higgs, S. Impact of climate change and other factors on emerging arbovirus diseases. *Transactions of the Royal Society of Tropical Medicine and Hygiene 103*, 2 (2009), 109.

Græsbøll, K., Bødker, R., Enøe, C., and Christiansen, L. E. Simulating spread of bluetongue virus by flying vectors between hosts on pasture. *Sci. Rep. 2*, 863 (2012).

Græsbøll, K., Sumner, T., Enøe, C., Christiansen, L. E., and Gubbins, S. A comparison of dynamics in two models for the spread of a vector borne disease. *Sci. Rep. ?*, Submitted (2012/2013).

Gubbins, S., Carpenter, S., Baylis, M., Wood, J. L. N., and Mellor, P. S. Assessing the risk of bluetongue to uk livestock: uncertainty and sensitivity analyses of a temperature-dependent model for the basic reproduction number. *Journal of The Royal Society Interface 5*, 20 (2008), 363–371.

Gubbins, S., Hartemink, N. A., Wilson, A. J., Moulin, V., Vonk Noordegraaf, C. A., van der Sluijs, M. T. W., de Smit, A. J., Sumner, T., and Klinkenberg, D. Scaling from challenge experiments to the field: Quantifying the impact of vaccination on the transmission of bluetongue virus serotype 8. *Preventive Veterinary Medicine 105* (2012).

Guis, H., Tran, A., de La Rocque, S., Baldet, T., Gerbier, G., Barrague, B., Biteau-Coroller, F., Roger, F., Viel, J. F., and Mauny, F. Use of high spatial resolution satellite imagery to characterize landscapes at risk for bluetongue. *Vet. Res. 38* (2007), 669–683.

Hartemink, N. Vector-borne diseases: the basic reproduction number R_0 and risk maps. http://igitur-archive.library.uu.nl/dissertations/20091126200122/-hartemink.pdf, 2009.

Hartemink, N. A., Purse, B. V., Meiswinkel, R., Brown, H. E., de Koeijer, A., Elbers, A. R. W., Boender, G. J., Rogers, D. J., and Heesterbeek, J. A. P. Mapping the basic reproduction number (R_0) for vector-borne diseases: A case study on bluetongue virus. *Epidemics 1*, 3 (2009), 153 – 161.

Hendrickx, G., Gilbert, M., Staubach, C., Elbers, A., Mintiens, K., Gerbier, G., and Ducheyne, E. A wind density model to quantify the airborne spread of Culicoides species during north-western Europe bluetongue

epidemic, 2006. *Preventive Veterinary Medicine 87*, 1-2 (2008), 162–181.

Hoffmann, B., Scheuch, M., Hoper, D., Jungblut, R., Holsteg, M., Schirrmeier, H., Eschbaumer, M., Goller, K. V., Wernike, K., Fischer, M., Breithaupt, A., Mettenleiter, T. C., and Beer, M. Novel orthobunyavirus in Cattle, Europe, 2011. *Emerging Infect. Dis. 18*, 3 (Mar 2012), 469–472.

Holmes, P. R., and Birley, M. H. An improved method for survival rate analysis from time series of haematophagous dipteran populations. *The Journal of Animal Ecology 56*, 2 (1987), 427–440.

Jennings, D. M., and Mellor, P. S. The vector potential of British Culicoides species for bluetongue virus. *Veterinary Microbiology 17*, 1 (1988), 1 – 10.

Jewell, C. P., Keeling, M. J., and Roberts, G. O. Predicting undetected infections during the 2007 foot-and-mouth disease outbreak. *Journal of The Royal Society Interface 6*, 41 (2009), 1145–1151.

Jones, A., Thomson, D., Hort, M., and Devenish, B. The U.K. met office's next-generation atmospheric dispersion model, NAME III. In *Air Pollution Modeling and Its Application XVII*, C. Borrego and A.-L. Norman, Eds. Springer US, 2007, pp. 580–589.

Kachitvichyanukul, V., and Schmeiser, B. W. Binomial random variate generation. *Communications of the ACM 31*, 2 (1988), 216–222.

Kampen, H., and Werner, D. Three years of bluetongue disease in central Europe with special reference to germany: what lessons can be learned? *Wiener Klinische Wochenschrift 122* (2010), 31–39.

Keeling, M. J. Models of foot-and-mouth disease. *Proc. Biol. Sci. 272*, 1569 (Jun 2005), 1195–1202.

Kirkeby, C., Bødker, R., Stockmarr, A., and Enøe, C. Association between land cover and Culicoides (diptera: Ceratopogonidae) breeding sites on four danish cattlefarms. *Entomologica Fennica 20*, 4 (2009), 228–232.

Kirkeby, C., Bødker, R., Stockmarr, A., Lind, P., and Heegaard, P. M. Quantifying dispersal of european culicoides (diptera: Ceratopogonidae) vectors between farms using a novel mark-release-recapture technique. *PloS one 8*, 4 (2013), e61269.

Kirkpatrick, S., Gelatt Jr, C. D., and Vecchi, M. P. Optimization by simulated annealing. *Science 220*, 4598 (1983), 671–680.

Lassen, S., Nielsen, S., and Kristensen, M. Identity and diversity of blood meal hosts of biting midges (diptera: Ceratopogonidae: Culicoides latreille) in denmark. *Parasites & Vectors 5* (2012), 143.

Lassen, S., Nielsen, S., Skovgård, H., and Kristensen, M. Molecular identification of bloodmeals from biting midges (diptera: Ceratopogonidae: Culicoides latreille) in denmark. *Parasitology research 108*, 4 (2011), 823–829.

Little, R. J. A., and Rubin, D. B. *Statistical analysis with missing data*, vol. 1. Wiley New York, 1987.

Lorca-Oró, C., Pujols, J., García-Bocanegra, I., Mentaberre, G., Granados, J., Solanes, D., Fandos, P., Galindo, I., Domingo, M., Lavín, S., et al. Protection of spanish ibex (capra pyrenaica) against bluetongue virus serotypes 1 and 8 in a subclinical experimental infection. *PloS one 7*, 5 (2012), e36380.

Maan, S., Maan, N. S., Nomikou, K., Batten, C., Antony, F., Belaganahalli, M. N., Samy, A. M., Reda, A. A., Al-Rashid, S. A., El Batel, M., et al. Novel bluetongue virus serotype from Kuwait. *Emerging infectious diseases 17*, 5 (2011), 886.

Maclachlan, N. J. Bluetongue: History, global epidemiology, and pathogenesis. *Preventive Veterinary Medicine 102*, 2 (2011), 107–111.

Madsen, H. *Time series analysis*. CRC Press, 2008.

Marangon, S., Cecchinato, M., and Capua, I. Use of vaccination in avian influenza control and eradication. *Zoonoses and Public Health 55*, 1 (2008), 65–72.

Meiswinkel, R., Baylis, M., and Labuschagne, K. Stabling and the protection of horses from Culicoides bolitinos (diptera: Ceratopogonidae), a recently identified vector of african horse sickness. *Bulletin of entomological research 90*, 06 (2000), 509–515.

Mellor, P., and Hamblin, C. African horse sickness. *Veterinary research 35*, 4 (2004), 445–466.

Mellor, P. S. Culicoides as potential orbivirus vectors in Europe. In *Bluetongue, African horse sickness and related orbiviruses (TE Walton & BI Osburn,*

eds). Proc. Second International Symposium, Paris (1991), pp. 17–21.

Mellor, P. S. Infection of the vectors and bluetongue epidemiology in Europe. *Veterinaria Italiana 40*, 3 (2004), 167.

Mellor, P. S., Boorman, J., and Baylis, M. Culicoides biting midges: their role as arbovirus vectors. *Annual review of entomology 45*, 1 (2000), 307–340.

Mellor, P. S., Carpenter, S., Harrup, L., Baylis, M., and Mertens, P. P. C. Bluetongue in Europe and the mediterranean basin: History of occurrence prior to 2006. *Preventive Veterinary Medicine 87*, 1–2 (2008), 4–20.

Mellor, P. S., and Wittmann, E. J. Bluetongue virus in the mediterranean basin 1998 - 2001. *The Veterinary Journal 164*, 1 (2002), 20–37.

Melville, L., Weir, R., Harmsen, M., Walsh, S., Hunt, N., and et al. Characteristics of naturally-occurring bluetongue viral infections of cattle. In *Bluetongue Disease in Southeast Asia and the Pacific.* (1996), T. St George and K. Peng, Eds., Canberra: ACIAR proceedings series, pp. 245–250.

Menzies, F. D., Burns, K., Fallows, J. G., McKeown, I. M., Gloster, J., Burgin, L., Courcier, E. A., McNamee, P. T., Robinson, J., Robinson, P. A., et al. Risk assessment, targeted surveillance and policy formulation: practical experiences with bluetongue. In *International Conference on Animal Health Surveillance (ICAHS), Lyon, France, 17-20 May, 2011.* (2011), no. 59/60, Association pour l'Étude de l'Épidémiologie des Maladies Animales, pp. 113–115.

Menzies, F. D., McCullough, S. J., McKeown, I. M., Forster, J. L., Jess, S., Batten, C., Murchie, A. K., Gloster, J., Fallows, J. G., Pelgrim, W., Mellor, P. S., and Oura, C. A. Evidence for transplacental and contact transmission of bluetongue virus in cattle. *Vet. Rec. 163*, 7 (Aug 2008), 203–209.

Muirhead-Thomson, R., et al. *Trap responses of flying insects. The influence of trap design on capture efficiency.* Academic Press Limited, 1991.

Mullens, B. A., Gerry, A. C., Lysyk, T. J., and Schmidtmann, E. T. Environmental effects on vector competence and virogenesis of bluetongue virus in Culicoides: interpreting laboratory data in a field context. *Vet. Ital. 40* (2004), 160–166.

Nielsen, S., Nielsen, B. O., Axelsen, J. A., and Fotel, F. Forurening af egeløkke lung og opformering af mitter. Tech. rep., Miljøstyrelsen, 1996.

Nielsen, S., Nielsen, B. O., Axelsen, J. A., and Fotel, F. Aktivitet af mitter på græsningsarealer ved egeløkke lung. Tech. rep., Miljøstyrelsen, 2003.

Ninio, C., Augot, D., Dufour, B., and Depaquit, J. Emergence of Culicoides obsoletus from indoor and outdoor breeding sites. *Veterinary Parasitology 183* (2011), 125–129.

Nowinszky, L. Nocturnal illumination and night flying insects. *Appl. Ecol. Environ. Res 2* (2004), 17–52.

Odetoyinbo, J. A. Preliminary investigation on the use of a light-trap for sampling malaria vectors in the gambia. *Bulletin of the World Health Organization 40*, 4 (1969), 547.

Ortega, M. D., Mellor, P. S., Rawlings, P., and Pro, M. J. The seasonal and geographical distribution of Culicoides imicola, C. pulicaris group and C. obsoletus group biting midges in central and southern Spain. *Arch. Virol. Suppl. 14* (1998), 85–91.

Paweska, J. T., Venter, G. J., and Mellor, P. S. Vector competence of south african Culicoides species for bluetongue virus serotype 1 (BTV-1) with special reference to the effect of temperature on the rate of virus replication in c. imicola and c. bolitinos. *Medical and Veterinary Entomology 16*, 1 (2002), 10–21.

Pinheiro, F. P., da Rosa, A. P. A. T., da Rosa, J. F. S. T., Ishak, Freita, R. B., R., GOMES, M., LeDuc, J., and Oliva, O. F. P. Oropouche virus, a review of clinical, epidemiological and ecological findings. *Am J Trop Med Hyg 30* (1981), 149–160.

Pioz, M., Guis, H., Crespin, L., Gay, E., Calavas, D., Durand, B., Abrial, D., and Ducrot, C. Why did bluetongue spread the way it did? environmental factors influencing the velocity of bluetongue virus serotype 8 epizootic wave in france. *PloS one 7*, 8 (2012), e43360.

Purse, B. V., Mellor, P. S., Rogers, D. ., Samuel, A. R., Mertens, P. P. C., and Baylis, M. Climate change and the recent emergence of bluetongue in Europe. *Nature Reviews Microbiology 3*, 2 (2005), 171–181.

Purse, B. V., Tatem, A. J., Caracappa, S., Rogers, D. J., Mellor, P. S., Baylis, M., and Torina, A. Modelling the distributions of Culicoides

bluetongue virus vectors in sicily in relation to satellite-derived climate variables. *Medical and Veterinary Entomology 18*, 2 (2004), 90–101.

Rasmussen, L. D., Kristensen, B., Kirkeby, C., Rasmussen, T. B., Belsham, G. J., and Bødker, R. Culicoides as vectors of schmallenberg virus [letter]. *Emerging Infectious Diseases* (2012).

Rigot, T., and Gilbert, M. Quantifying the spatial dependence of Culicoides midge samples collected by onderstepoort-type blacklight traps: an experimental approach to infer the range of attraction of light traps. *Medical and Veterinary Entomology 26*, 2 (2012), 152–161.

Rodriguez, M., Hooghuis, H., and Castaño, M. African horse sickness in Spain. *Veterinary Microbiology 33*, 1-4 (1992), 129–142.

Sanders, C. J., Gubbins, S., Mellor, P. S., Barber, J., Golding, N., Harrup, L. E., and Carpenter, S. T. Investigation of diel activity of Culicoides biting midges (diptera: Ceratopogonidae) in the united kingdom by using a vehicle-mounted trap. *Journal of medical entomology 49*, 3 (2012), 757–765.

Sanders, C. J., Shortall, C. R., Gubbins, S., Burgin, L., Gloster, J., Harrington, R., Reynolds, D. R., Mellor, P. S., and Carpenter, S. Influence of season and meteorological parameters on flight activity of Culicoides biting midges. *Journal of Applied Ecology 48*, 6 (2011), 1355–1364.

Savini, G., Goffredo, M., Monaco, F., Di Gennaro, A., Cafiero, M. A., Baldi, L., De Santis, P., Meiswinkel, R., and Caporale, V. Bluetongue virus isolations from midges belonging to the Obsoletus complex (Culicoides, diptera: Ceratopogonidae) in italy. *Veterinary Record 157*, 5 (2005), 133–139.

Savini, G., Hamers, C., Conte, A., Migliaccio, P., Bonfini, B., Teodori, L., Di Ventura, M., Hudelet, P., Schumacher, C., and Caporale, V. Assessment of efficacy of a bivalent btv-2 and btv-4 inactivated vaccine by vaccination and challenge in cattle. *Veterinary microbiology 133*, 1 (2009), 1–8.

Savini, G., MacLachlan, N. J., Sánchez-Vizcaino, J.-M., and Zientara, S. Vaccines against bluetongue in Europe. *Comparative Immunology, Microbiology and Infectious Diseases 31*, 2–3 (2008), 101–120.

Sedda, L., Brown, H. E., Purse, B. V., Burgin, L., Gloster, J., and Rogers, D. J. A new algorithm quantifies the roles of wind and midge flight activity in the bluetongue epizootic in northwest Europe. *Proc. Biol. Sci. 279*,

1737 (Jun 2012), 2354–2362.

Sellers, R., and Pedgley, D. Possible windborne spread to western Turkey of bluetongue virus in 1977 and of akabane virus in 1979. *Journal of hygiene 95*, 1 (1985), 149–158.

Silbermayr, K., Hacklände, K., Doscher, C., Koefer, J., and Fuchs, K. A spatial assessment of Culicoides spp. distribution and bluetongue disease risk areas in Austria. *Berliner und Münchener tierärztliche Wochenschrift 124*, 5-6 (2011), 228.

Silver, J. *Mosquito ecology: field sampling methods*. Springer Verlag, 2007.

Spreull, J. Malarial catarrhal fever (bluetongue) of sheep in South Africa. *J Comp Pathol Ther 18* (1905), 321–337.

Stott, J., Oberst, R., Channell, M., and Osburn, B. Genome segment reassortment between two serotypes of bluetongue virus in a natural host. *Journal of virology 61*, 9 (1987), 2670–2674.

Szmaragd, C., Wilson, A. J., Carpenter, S., Wood, J. L. N., Mellor, P. S., and Gubbins, S. A modeling framework to describe the transmission of bluetongue virus within and between farms in great britain. *PLoS ONE 4*, 11 (11 2009), e7741.

Szmaragd, C., Wilson, A. J., Carpenter, S., Wood, J. L. N., Mellor, P. S., and Gubbins, S. The spread of bluetongue virus serotype 8 in great britain and its control by vaccination. *PloS one 5*, 2 (2010), e9353.

Tabachnick, W. J. Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world. *Journal of Experimental Biology 213*, 6 (2010), 946–954.

Tabachnick, W. J., et al. Culicoides and the global epidemiology of bluetongue virus infection. *Veterinaria Italiana 40*, 3 (2004), 145.

Takken, W., Verhulst, N., Scholte, E. J., Jacobs, F., Jongema, Y., and van Lammeren, R. The phenology and population dynamics of Culicoides spp. in different ecosystems in the netherlands. *Preventive veterinary medicine 87*, 1 (2008), 41–54.

Truxa, C., and Fiedler, K. Attraction to light–from how far do moths (lepidoptera) return to weak artificial sources of light? *Eur. J. Entomol 109*, 1 (2012), 77–84.

Turner, J., Bowers, R. G., and Baylis, M. Modelling bluetongue virus transmission between farms using animal and vector movements. *Sci. Rep. 2* (03 2012).

Velthuis, A. G. J., Saatkamp, H. W., Mourits, M. C. M., de Koeijer, A. A., and Elbers, A. R. W. Financial consequences of the dutch bluetongue serotype 8 epidemics of 2006 and 2007. *Preventive Veterinary Medicine 93*, 4 (2010), 294–304.

Venter, G., Hermanides, K., Boikanyo, S., Majatladi, D., and Morey, L. The effect of light trap height on the numbers of Culicoides midges collected under field conditions in south africa. *Veterinary parasitology 166*, 3 (2009), 343–345.

Venter, G. J., Labuschagne, K., Hermanides, K. G., Boikanyo, S. N. B., Majatladi, D. M., and Morey, L. Comparison of the efficiency of five suction light traps under field conditions in south africa for the collection of Culicoides species. *Veterinary parasitology 166*, 3 (2009), 299–307.

Venter, G. J., Majatladi, D. M., Labuschagne, K., Boikanyo, S. N. B., and Morey, L. The attraction range of the Onderstepoort 220v light trap for Culicoides biting midges as determined under south african field conditions. *Veterinary Parasitology in press* (2012).

Viswanathan, G. M., Buldyrev, S. V., Havlin, S., Da Luz, M. G. E., Raposo, E. P., and Stanley, H. E. Optimizing the success of random searches. *Nature 401*, 6756 (1999), 911–914.

Wall, C., and Perry, J. Range of action of moth sex-attractant sources. *Entomologia experimentalis et applicata 44*, 1 (1987), 5–14.

Wilson, A., and Mellor, P. Bluetongue in Europe: vectors, epidemiology and climate change. *Parasitology Research 103* (2008), 69–77.

Wilson, A. J., and Mellor, P. S. Bluetongue in Europe: past, present and future. *Philos. Trans. R. Soc. Lond., B, Biol. Sci. 364* (Sep 2009), 2669–2681.

Yadin, H., Brenner, J., Bumbrov, V., Oved, Z., Stram, Y., Klement, E., Perl, S., Anthony, S., Maan, S., Batten, C., et al. Epizootic haemorrhagic disease virus type 7 infection in cattle in Israel. *Veterinary Record 162*, 2 (2008), 53–56.

Yanase, T., Maeda, K., Kato, T., Nyuta, S., Kamata, H., Yamakawa, M., and Tsuda, T. The resurgence of shamonda virus, an african simbu group

virus of the genus orthobunyavirus, in japan. *Archives of virology 150*, 2 (2005), 361–369.